# UC Berkeley UC Berkeley Electronic Theses and Dissertations

## Title

Before the Eye Moves: Remapping, Visual Stability and Perisaccadic Perception

Permalink https://escholarship.org/uc/item/0j45z86n

Author Wolfe, Benjamin Arthur

Publication Date 2015

Peer reviewed|Thesis/dissertation

Before the Eye Moves: Remapping, Visual Stability and Perisaccadic Perception

by

## Benjamin Arthur Wolfe

A dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy

in

Psychology

in

the Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor David Whitney, Chair Professor Sonia Bishop Professor Dennis M. Levi

Spring 2015

#### Abstract

### Before the Eye Moves: Remapping, Visual Stability and Perisaccadic Perception

by

#### Benjamin Arthur Wolfe

#### Doctor of Philosophy in Psychology

#### Professor David Whitney, Chair

Our perception is of a stable visual world, yet we make several saccadic eye movements per second – to say nothing of the motion of objects in the world or our own motion through the physical world. Even standing on a stage, looking out at an audience, we have the perception of the crowd in detail, in spite of our less than accurate peripheral input. Standing on this stage, it is a trivial matter to make a series of saccades to the faces of various audience members to find one particular person – how can we do this, since visual crowding makes these faces unidentifiable? How much information is available from peripheral vision prior to a saccade? While making saccades to individual members of the crowd is useful for identifying them, what about gaining a sense of the audience as a whole? We can perceive the ensemble emotion of the audience without examining each and every member, but is this solely possible with peripheral information, the details of which we cannot otherwise access, or does it require foveal information?

While a great deal of information is available from peripheral vision, identifying individual objects is difficult due to visual crowding; we solve this problem by simply making a saccade to the object we wish to identify. The ability to saccade to a crowded object implied that a saccade to a crowded object might reduce crowding before the object was foveated, essentially unlocking peripheral information that is otherwise inaccessible. We performed a series of experiments where subjects made saccades to crowded faces and were not permitted to foveate them; we found that crowding was diminished presaccadically and that this was more effective than covert attention alone. Our crowding result shows that saccade planning can diminish crowding, but what information becomes available to conscious vision? To determine this, we performed a series of experiments to determine whether saccadic remapping acquired object representations or merely constituent features from saccade targets. We found that saccadic remapping was object-selective and tightly tuned to the target of the saccade, suggesting that, prior to a saccade, detailed information is made available to conscious vision. However, there simply is not time to take advantage of this unlocking for each face in a crowd - can the visual system use this inaccessible information to generate a percept of the group as a whole – an ensemble code – and is foveation of individuals required? We occluded foveal input and determined that it was unnecessary for ensemble perception of emotion. In all, these three studies demonstrate that peripheral input is remarkably detailed and that the visual system can use this information to both facilitate identification of individual group members as part of saccade planning, and to provide a useful assessment of the crowd as a whole.

Dedication

This doctoral dissertation is dedicated to my fiancé, Anna Kosovicheva, without whom achieving my Doctorate would have been impossible.

## Table of Contents

Dedication	i
Acknowledgements	iii
Chapter 1 : Introduction	. 1
Chapter 2 : Facilitating Recognition of Crowded Faces with Presaccadic Attention	. 2
Chapter 3 : Saccadic Remapping of Object-Selective Information	21
Chapter 4 : Foveal Input is not Required for Perception of Ensemble Facial Expression 3	38
Chapter 5 : Conclusion	55
References	56

#### Acknowledgements

This dissertation would have been impossible without the members of the Whitney Lab, past and present, who have contributed to the research described herein in innumerable ways in the last five years. These include:

Alice Albrecht Yang Bai Wesley Chaney Kelly Chang Jason Fischer Anna Kosovicheva Alina Liberman Allison Yamanashi Leib Mauro Manassi Gerrit Maus Tim Sweeny Santani Teng David Whitney Katherine Wood Nicole Wurnitsch Ye Xia Kathy Zhang

Particular thanks, of course, to my fiancé and collaborator, Anna Kosovicheva

Additional thanks to Funranium Labs and Phillip Broughton for providing awe-inspiring amounts of astonishingly powerful coffee to further Science.

Funding for the research described in this Dissertation was generously provided by the National Science Foundation and the National Institutes of Health

#### Chapter 1 : Introduction

Imagine that you are standing on the stage of Wheeler Auditorium, lecturing to six hundred students – how does your visual system generate a stable perception of the world for you, much less allow you to both make eye movements to individual students in the crowd, while simultaneously maintaining an ensemble percept of the crowd's overall emotional state? Given the density of faces looking at you on the stage, they are crowded from your awareness – you know that they are there, but you do not particularly know who is where. How can you identify an individual within the crowd? How can you assess the crowd as a whole?

The easy solution to the first problem is to just make a saccade to whoever you need to look at – but how can you do this when you cannot identify individuals? The visual periphery is detailed, so can planning a saccade diminish crowding before you have even moved your eyes? The individual identities are crowded from your awareness, but are they accessible to unconscious processes within your visual system? Chapter 2 addresses this question through two experiments examining the ability of saccades to reduce crowding between emotional faces, and demonstrates that the act of making an accurate saccade to a crowded face renders that face more identifiable than it would be through covert examination alone, even though that face is never foveated. However, this result simply indicates that crowding is reduced through saccade planning; it does not and cannot speak to what peripheral information is made available to conscious vision prior to the saccade. Towards this end, Chapter 3 is a detailed investigation of the limits of saccadic remapping, the ability of the visual system to respond to the target of an impending saccade before the eye has moved. In the chapter, we show that saccadic remapping is a remarkably capable process which acquires object-specific information rather than individual features, and that it does so in a spatially-specific manner. Together, Chapters 2 and 3 show that the visual system captures a remarkable amount of detail in the periphery, even though this information is usually inaccessible and that it is made available to conscious vision immediately prior to a saccade.

However, while peripheral vision is remarkably detailed, its contents are not usually available to conscious vision outside the circumstances we have just discussed. To return to our scene, how can we gain a sense of the crowd as a whole? Peripheral input is detailed, but we do not have conscious access to a representation of each and every member of the crowd – can we generate an ensemble percept of the crowd's emotion without making an eye movement to each and every member? Chapter 4 addresses this question, and also asks if generating a percept of the crowd's average emotion requires foveal input at all. Foveal input does not seem to be required to judge the average of a group – there is, in fact, no effect of foveal information on subjects' ability to do so. Overall, the research discussed in this dissertation shows that peripheral input is detailed, reflecting the identity of individuals in the crowd, and that this information can be accessed for individual objects via saccade planning, or globally averaged for an ensemble percept without being unlocked by saccade planning.

#### Chapter 2 : Facilitating Recognition of Crowded Faces with Presaccadic Attention

Observers make several saccades per second to foveate objects in the world, since objects near other objects are often crowded from our awareness. We can see these objects and we have a sense of where they are in visual space, but we cannot identify them without saccading to and foveating them. This is the underlying assumption of much of our visual system – that, in order to better identify an object, we must make a saccade to it – and until recently, the focus had been on how do we make a saccade, rather than what information do we acquire in the process. Saccadic eye movements have been studied for over a century (beginning with Javal, 1878, who coined the term; translated by (Huey, 1908); see (Kowler, 2011) for an extensive review of the current state of the art. However, it is only relatively recently, that we have started to ask about what information is acquired prior to the start of a saccade. In particular, we make saccades to crowded objects in order to identify them.

Natural scenes are rich with objects to the point of being cluttered, and this results in visual crowding, a central problem of conscious vision. Visual crowding is the inability to identify an object in the periphery when it is surrounded by other stimuli (Bouma, 1970) and can be operationally defined as the change in the ability to identify an object in the periphery as a result of proximate objects (flankers) in space or time, often phenomenologically reported as a "jumbled" percept of the object and its proximate flankers. Crowding is not a problem of detection; an object or feature is perceived to be present at a location, but it is unidentifiable and its features are jumbled (see Korte, 1923 as translated in (Pelli, Palomares, & Majaj, 2004; Strasburger, Rentschler, & Juttner, 2011). Crowding has been studied extensively (Levi, 2008; Pelli, 2008; Whitney & Levi, 2011) with the vast majority of this research having been performed with comparatively simple combinations of features, or letters (Balas, Nakano, & Rosenholtz, 2009; Chung, Li, & Levi, 2007; Grainger, Tydgat, & Isselé, 2010; Jeon, Hamid, Maurer, & Lewis, 2010; Schotter, Angele, & Rayner, 2012). However, crowding also occurs selectively between objects (Chakravarthi & Cavanagh, 2009; Greenwood, Bex, & Dakin, 2010; Wallace & Tjan, 2011) and faces (Farzin, Rivera, & Whitney, 2009; Fischer & Whitney, 2011; Louie, Bressler, & Whitney, 2007). In addition, considerable research has been done on the impact of flanker configuration on the strength and reduction of crowding (Chakravarthi & Pelli, 2011; Kowler, 2011; Livne & Sagi, 2007; 2010; Malania, Herzog, & Westheimer, 2007; Manassi, Sayim, & Herzog, 2012; Sayim, Westheimer, & Herzog, 2008; 2010). Crowding often ceases to be a problem once the crowded object has been foveated (Pelli & Tillman, 2008), yet the perceptual consequences of information acquired prior to the saccade on crowding have barely been investigated.

Making a saccade to an uncrowded target has been shown to facilitate identification (Remington, 1980) and more recent work has shown that presaccadic attention is key to both the saccade planning process and post-saccade identification of the target object (Kowler, Anderson, Dosher, & Blaser, 1995; Schneider & Deubel, 1995). However, these studies of presaccadic attention have been limited to relatively simple stimuli (e.g., the "E" and "mirror-E" stimuli used in (Schneider & Deubel, 1995) and (Deubel & Schneider, 1996). While presaccadic attention is undoubtedly an important mechanism

for presaccadic information acquisition, recent work by Wurtz and colleagues (Wurtz, 2008; Wurtz, Joiner, & Berman, 2011), building on the extensive corollary discharge literature, has suggested that the corollary discharge prior to a saccade may act as a trigger for non-attentional facilitation in the same time window. This facilitation would be complementary to presaccadic attention as described by Schneider, Deubel and Kowler respectively, and suggests that there may be alternate mechanisms by which the visual system acquires detailed information about a saccade target prior to a saccade. It is therefore reasonable that saccade planning might reduce crowding.

Certainly, making a saccade to a crowded object and foveating that object thereafter breaks crowding and allows for the crowded object to be easily identified (Pelli & Tillman, 2008) – but, given the results of recent studies (Deubel & Schneider, 1996; Kowler et al., 1995; Schneider & Deubel, 1995) examining presaccadic attention, we wondered if presaccadically acquired information could reduce crowding on its own. Several recent studies have suggested that saccadic eye movements are accompanied by enhanced attentional resolution (Baldauf, Wolf, & Deubel, 2006; Deubel, 1995; Deubel & Schneider, 1996), and reduced critical spacing in crowding of features or letters (Harrison, Mattingley, & Remington, 2013a).

As we have discussed, crowding has a well-known adverse impact on identification (as reviewed by (Pelli, 2008; Pelli et al., 2004; Whitney & Levi, 2011), but it does not impact saccadic accuracy to crowded targets (Vlaskamp & Hooge, 2006), suggesting that sufficient position information is available for accurate saccade planning to crowded targets. More recently, Harrison and colleagues (Harrison, Mattingley, & Remington, 2013a) showed that saccades to crowded Gabors acted to reduce the effects of crowding, with the orientation of the Gabors becoming more identifiable by subjects when the Gabors were presented in the 50ms immediately prior to saccade onset. While Harrison and colleagues found a reduction in crowding from presaccadic input alone, their results, while suggestive, do not tell us if the same effect might be found with crowded faces. Additional striking work by Harrison and colleagues (Harrison, Retell, Remington, & Mattingley, 2013b), using a set of letters as stimuli, has shown that visual features can be presaccadically remapped to induce crowding, indicating that saccadic remapping precedes at least one stage of crowding. Incidentally, their work also confirmed the earlier results of Vlaskamp and Hooge with different stimuli, showing that crowding does not impact saccade accuracy.

While earlier work (Deubel & Schneider, 1996; Kowler et al., 1995) demonstrated the crucial role presaccadic attention plays in the saccade planning process, they did not investigate the perceptual consequences of the information that is acquired prior to the saccade. Most importantly for our purposes, the studies cited above on presaccadic attention did not explicitly measure crowding, as noted by Harrison and colleagues (2013a; c.f.,(Harrison, Mattingley, & Remington, 2013a; van Koningsbruggen & Buonocore, 2013). Moreover, all the work done to date on saccadic amelioration of crowding has used Gabors as stimuli, which, while revealing, are not as complex as stimuli that we commonly saccade to, namely, faces. It remains unclear if the results of Harrison and colleagues (Harrison, Mattingley, & Remington, 2013a) extend to more

complex and naturalistic face stimuli (Maurer, Grand, & Mondloch, 2002; McKone, Kanwisher, & Duchaine, 2007).

While Harrison and colleagues (Harrison, Mattingley, & Remington, 2013a; Harrison, Retell, Remington, & Mattingley, 2013b) found a modulation of crowding from presaccadic input alone, it remains unclear if there results extend to other visually complex stimuli such as faces. Thus, the experiments of Harrison and colleagues prompt another question: does face recognition in crowded scenes also benefit from saccades? In the following experiments, we asked if saccades to crowded faces could diminish the effects of crowding, making the faces more identifiable. Gabors and faces are processed at different levels of analysis, faces selectively crowd each other (Farzin et al., 2009; Haberman & Whitney, 2009; Louie et al., 2007), and Gabor (or letter) crowding does not account for the crowding of faces (Whitney & Levi, 2011). Therefore, although saccades might mitigate crowding of features, they may or may not modulate crowding of high-level objects such as faces. In this study, we tested whether presaccadic processes might improve recognition of crowded faces.

#### Methods

## **Display setup**

The experiments were performed using Matlab 2010a (Mathworks; Natick, MA) the Psychtoolbox (Brainard, 1997; Pelli, 1997) and the Eyelink Toolbox (Cornelissen, Peters, & Palmer, 2002) on a Mac Mini (Apple; Cupertino, CA). Stimuli were displayed on a 47 cm Samsung cathode ray tube display 57 cm from the subject; the display resolution was set to a resolution of 1024 x 768 at 60 Hz in all experiments.

## Subjects

Four subjects (two female; mean age 24) with extensive experience with psychophysical tasks, including the first author (BW), participated in the first experiment. With the exception of BW, all other subjects in the first experiment were naïve to the purpose of the experiment; and had normal or corrected to normal vision. Two subjects (BW and one naïve subject) participated in the control experiment. All subjects provided informed consent as required by the IRB at the University of California, Berkeley in accordance with the Declaration of Helsinki. All subjects received a minimum of 1000 trials of practice prior to any data collection.



Figure 2.1: (a) Schematic of trial structure in fixation and saccade conditions, with details of the stimuli. Note that the screen-by-screen schematic shows the crowded array in the upper right visual field; in all experiments, the crowded array was randomly presented in the upper and lower right visual field on a trial-by-trial basis. The size of the target face and flankers are exaggerated for illustration; all faces in all experiments were 2°x3°. The arrow in the saccade trial illustrates the subject's saccade to the target face. (b) Crowded stimulus array used in Experiment 1 with two inverted flanking faces. (c) Illustration of crowded stimulus array as used in Experiment 2, showing two of five possible target-flanker spacings with four inverted flanking faces. (d) The less disgusted and more disgusted target faces used in both experiments.

## Stimuli

All stimuli were faces originally from Ekman's Pictures of Visual Affect (POFA as used by (Fischer & Whitney, 2011); subjects were trained on identifying two emotional faces (target stimuli) prior to any data collection (Figure 2.1d). All faces were portraits of a single Caucasian female, initially morphed between a neutral and a disgusted expression across 48 intermediate, computer-generated faces, corrected to have the same mean luminance. All experiments used a subset of this space, from a moderately disgusted face (#25; halfway through the morph space) to a fully disgusted face (#50; see Figure 2.1d for both). All flanking stimuli were randomly sampled around the median point (e.g., halfway between "moderately disgusted" and "disgusted") of the two target faces. In all experiments, flankers were inverted (as shown in Figure 2.1b and 2.1c) to avoid subject confusion in saccade targeting as well as to reduce possible flanker substitutions and perceptual pooling (Fischer & Whitney, 2011), and to reduce incorrectly directed saccades. Given a constant flanker-target spacing, flanker inversion reduces crowding, but does not eliminate it (Louie et al., 2007); by reducing the target-flanker spacing (Figure 2.1), we maximized crowding.

In all experiments, the target face and flankers were presented in the right visual field, 10° from the fixation cross, with the flankers in a radial orientation (Figure 2.1b, 2.1c) around the upright target face. Each face subtended a 3° x 2° region on the display. Note that in Figure 2.1a, the crowded faces are shown in the upper right quadrant of the screen; the presentation location was randomized between the upper right and lower right quadrants on a trial by trial basis. In Experiment 2 (Figure 2.1c), we added two additional inverted flankers to the array to maximize task difficulty, one above and to the left of the target face and a second below and to the right of the target face to better assess the impact of inverted faces as flankers. Target-flanker spacing (center to center) remained constant at 3° for all trials in Experiment 1 and was varied on a trial-by-trial basis in Experiment 2. Experiment 2 is solely a control for the presence of crowding with face stimuli, since target-flanker spacing remained constant at 3° in Experiment 1. In Experiment 2, target-flanker spacing varied on a trial-by-trial basis from 2.5° to 4.5° in the saccade trials and 3° to 5° in the fixation trials in steps of 0.5°.

#### **Trial sequence**

Each trial, in all two experiments, began with a red or green fixation cross for 1400-1500 ms (randomized) on a light grey background (39.68 cd/m<sup>2</sup>). The fixation cross was onscreen at all times during the trial (Figure 2.1a). All stimuli were presented gaze-contingently with eye position monitored throughout the trial. If, at the start of the trial, the fixation cross was green, it signaled a saccade trial, red signaled a fixation trial. Subjects were then shown a valid spatial cue (a  $0.25^{\circ}$  dark grey square) for 150 ms at the location where the target face would appear. Once the cue appeared, the fixation cross became dark grey and remained so for the remainder of the trial. The cue was then removed from the screen and was immediately followed by the target face and its inverted flankers (Figure 2.1b and 2.1c for detail) centered on the location where the cue had appeared for a maximum of 200 ms. The location of the crowded stimulus array (the target face and the two flankers) was jittered by up to  $\pm 1^{\circ}$  of visual angle in the X and Y dimensions on a trial-by-trial basis to avoid stereotyped saccades. In the fixation

condition, the stimulus was onscreen for the full 200 ms. In the saccade condition, the stimulus was onscreen for 200 ms or until the saccade was initiated, whichever came first. Eye position was monitored online at 1000 Hz (see Eyetracking, 2.6), and in all saccade trials, any deviation of the subject's point of gaze greater than 0.5° was treated as the start of a saccade and resulted in the target face being immediately removed from the display. As a result of this conservative ocular motion threshold, the faces remained onscreen for 70 ms, on average, during the saccade trials and reflects a certain degree of saccade planning during the cue period. Gaze-contingent control was maintained throughout the experiment; any deviation from fixation in fixation trials greater than 0.5° initiated the removal of the stimulus, and the trial being redone at the end of the block.

Once the stimulus was removed, subjects made two responses using the keyboard; first, a 2AFC identification response (was the target presented face A [less disgusted] or face B [more disgusted], distinguishing between the two faces shown in Figure 2.1d) and second, a 5AFC confidence rating, with a rating of 1 indicating a total lack of confidence in their identification response and a rating of 5 indicating total confidence in their identification response on that trial. The second keyboard response triggered the next trial after a 2 second pause to allow subjects to move their eyes back to the start location (fixation cross). Collecting this additional 5AFC response wherein subjects rated their confidence in their own responses allowed us to better probe the effect of the saccade; subjects reported that they felt more confident in their own responses when making a saccade than when fixating.

## **Block sequence**

In Experiment 1, subjects participated in five runs of 216 trials each; each run consisted of eight 27-trial blocks which alternated between the saccade or fixation conditions. In Experiment 2, subjects participated in ten runs of 240 trials each, divided into eight 30-trial blocks which alternated between the saccade and fixation conditions.

#### Eyetracking

Eyetracking was performed using an Eyelink 1000 (SR Research; Mississauga, Ontario, Canada) with a level desktop camera; data was recorded monocularly (right eye for all subjects) at 1000 Hz and saccade analysis was performed offline using the Eyelink parser. A saccade was defined as the first time point at which the velocity exceeded 30°/s and the acceleration exceeded 8000°/s<sup>2</sup>. In addition, a motion threshold was used to delay the start of each saccade until the eye had moved at least 0.15°. All subjects were stabilized on a chinrest during all experiments. Subjects were calibrated using a standard 9-point grid.

Subjects' eye movements were recorded at all times during the experiment, and eye movements were only permitted in the saccade blocks during a specified response window after the crowded face was presented. Rather than attempt to parse saccades in real time, which would have introduced an unacceptable delay, raw gaze position was continually monitored over the realtime link between the eyetracking computer and the stimulus computer, and any deviation in eye position greater than 0.5° in any direction was treated as the beginning of a saccade. In the saccade trials, any deviation in excess of

0.5° resulted in the stimuli being immediately removed from the screen to prevent inadvertent foveation of the stimuli. Subjects were required to maintain accurate fixation during fixation trials, and we used the same criteria as in saccade trials to prevent inadvertent examination of the target faces. Eye movements at any other time, or during the fixation blocks, resulted in the trial being aborted, the screen going red for 2 s and the subject repeating the trial in random order at the end of the block.

#### Analysis

Behavioral and eyetracking data were analyzed offline using custom Matlab scripts and S-R Research's EDFMEX tool. All trials in the saccade blocks were filtered by landing location; subjects were required to land their first saccade within  $\pm 1.5^{\circ}$  of the center of the upright target face, and were otherwise discarded. Fixation trials were automatically discarded and rerun, as described in the previous section, if the subject's eye moved more than 0.5° at any time during a fixation trial. All data was collapsed across the upper and lower visual field presentation locations.

The confidence ratings and responses for each trial were used to calculate two Receiver Operating Characteristic (ROC) curves, one for the fixation condition and one for the saccade condition. The data points used to plot each ROC curve were calculated using the following procedure from (Murdock, 1965). Within each eye movement condition (saccade or fixation), we sorted each trial based on which face was presented (the less disgusted or more disgusted face) and then further sorted them based on the accuracy of subject's 2AFC identification responses and their 5AFC confidence ratings. The combination of the subject's 2AFC response and the 5-point rating of their confidence in their response together gave us a measure on each trial of how confident the subject was on a scale from 1-10 that they were presented with a more disgusted face. For example, if a subject responded that they were shown a "less disgusted" face and were highly confident in their response (i.e., gave a rating of 5), that meant that they were certain that they did not see a "more disgusted" face. In other words, the subject would be unconfident that they saw a more disgusted face, and this would be assigned a rating of 1 on the 1-10 scale. On the other hand, if the subject responded that they were shown a more disgusted face and were confident in their response (i.e., gave a rating of 5), they were confident that they saw a disgusted face, and this response would be assigned a rating of 10 on the 1-10 scale. Doing this across our entire set of trials within a given eye movement condition allowed us to reclassify our original 2AFC and 5AFC responses on a 10-point scale for each stimulus condition. We then calculated the number of responses for each rating (1 through 10) within each stimulus condition (disgusted or less disgusted) separately. We then calculated the proportion of the total number of trials within each stimulus condition that fell into each bin.

Using these proportions, we generated a set of cumulative conditional probabilities for each stimulus condition. In essence, for a given rating on our 10-point scale, we asked what proportion of responses would be encompassed by a criterion (for correctly detecting a disgusted face) set at a given point on the scale or greater. So, the cumulative conditional probability for a rating of 1 on this scale encompasses all of our transformed responses, while the cumulative conditional probability for a rating of 5 is considerably

less than that, reflecting the more stringent criterion. In each case, we can calculate hits (the cumulative probability for a given criterion when the subject was shown the disgusted face and correctly identified it) and false alarms (the cumulative probability when the subject was shown the less disgusted face and misidentified it as the disgusted face). The entire set of hit rates and false alarm rates for each criterion form the points of our ROC curve. We then calculated the area under curve using the trapezoidal integration function in Matlab. Significance was tested with Z-tests, the Z-scores were converted to p-values using the standard normal distribution on the area under curve (Hanley & McNeil, 1982). Standard error for area under curve was estimated using the method of (Hanley & McNeil, 1983).

## Results

## **Experiment 1**

Experiment 1 tested whether presaccadic information facilitated identification of a crowded emotional face in the absence of foveating the face; as described in Methods, target-flanker spacing was held constant at 3° throughout the experiment. The data were analyzed using a  $\pm 1.5^{\circ}$  window centered on the target face for saccade trials, ensuring that the saccades in question were well-localized to the crowded face and that the saccade on a given trial had not landed at the location where an inverted flanker had been present prior to the saccade.



Figure 2.2: Results from Experiment 1. (a-d): Receiver Operating Characteristic (ROC) curves from all four subjects showing performance in fixation and saccade conditions (significantly different; p<.01; all 4 subjects show significant differences individually) with saccade landings restricted to  $\pm 1.5^{\circ}$  from the center of the target face. Three of the four subjects (a-c) show significant performance in the fixation condition (comparison versus AUC of .5; p<.01). (e): ROC curves for all subjects (n=4) collapsed; saccade landings restricted as in (a-d). (f): Average area under curve for fixation (f; left bar) and saccade conditions (f; right). Trials on which the saccades landed up to  $\pm 1.5^{\circ}$  from the center of the face showed the greatest improvement in performance; the effect was reduced with less accurate saccades (landing  $\pm 1.5$ -3.0° from the center of the face). The area under curve in the fixation represents a baseline level of performance on the task. A Z-test was performed and the Z-score converted to a p-value using the standard normal distribution (Hanley & McNeil, 1982), and was then used to compare AUCs between the fixation condition and difference saccade accuracy windows; asterisks indicate significance at p < 0.05. Standard errors were estimated using the procedure of (Hanley & McNeil, 1983).

As a group (Figure 2.2e) and as individuals (Figure 2.2a-d), subjects found the identification task challenging, but they were able to perform the identification task above chance in the fixation condition (group AUC, 0.645 for fixation; p < 0.0001; twotailed Z-test converted to p-value; per the procedure of (Hanley & McNeil, 1982; Harrison, Mattingley, & Remington, 2013a), and subjects achieved significantly better performance in the saccade condition (group AUC, 0.735, p < 0.0001); the AUC for the saccade and fixation conditions were significantly different (p < 0.0001). Three out of four subjects, individually, were significantly above chance in the fixation condition (p<.01), indicating their basic competence at the task, and all subjects individually showed significant differences between the saccade and fixation conditions (Figure 2.2ad; all p<.01). We performed additional analyses (Figure 2.2d) excluding the central region previously analyzed  $(1.5^{\circ})$  and found similar results for the region of  $1.5-3^{\circ}$  from the center of the crowded face (saccade AUC, 0.703, p < 0.0001); grossly inaccurate landings (greater than 3°) did not show any significant difference between saccade and fixation performance at the group level. On average, 51% of saccades landed within ±1.5° of the center of the upright target face; an additional 42% landed within  $\pm 1.5$ -3°.

#### **Experiment 2**

To validate our procedure in Experiment 1, we performed a control experiment with two subjects to verify that subjects' reduction in performance in the fixation task was due to crowding rather than to other factors, such as limits on visual acuity. To do this we systematically varied the target-flanker spacing in Experiment 2. Observing a decrease in performance with a decrease in target-flanker spacing, and an increase in performance with an increase in target-flanker spacing would indicate that crowding did, in fact, occur with the stimulus arrangement we used in Experiment 1. In Experiment 2, we added two additional inverted flankers (Figure 2.1c) and used two offset ranges of target-flanker spacings, from 2.5° to 4.5° in the saccade trials, and 3° to 5° in the fixation trials, as well as adding a baseline no-flanker stimulus in both conditions as a point of comparison.



Figure 2.3: Averaged AUCs for Experiment 2, showing performance across a range of target-flanker spacings (2.5-4.5° in the saccade condition; 3-4.5° in the fixation condition) and a leftward shift in the saccade condition relative to the fixation baseline, suggesting that the saccade itself may diminish crowding. The data were fit to a logistic curve using the least-squares method in Matlab. Data in the no-flanker condition (NF, far right) is included as a baseline for comparison. Experiment 2 was performed with four flankers, as shown above the legend, to maximize the effects of crowding. A Z-test was performed and the Z-score converted to a p-value using the standard normal distribution (Hanley & McNeil, 1982), and was then used to compare AUCs between the fixation condition and difference saccade accuracy windows; general significance described in the text (Section 3.2). Standard errors were estimated using the procedure of (Hanley & McNeil, 1983).

In Experiment 2, we found that subjects showed equal performance with spacings in excess of 4° when making saccades and 4.5° when fixating; the average AUCs in Figure 2.3 show a decrease in performance with a decrease in target-flanker spacing, indicating that crowding impeded identification of the target face. A series of pairwise comparisons across the saccade and fixation conditions showed a significant difference in performance in the 3° target-flanker spacing condition (p<.001; Bonferroni-corrected  $\alpha = .01$ ); comparisons in the 3.5°, 4°, 4.5° and no-flanker conditions were all not significant (p>.01). We do find a significant overall effect of target-flanker spacing on performance; when we compared the closest spacing in the fixation condition  $(3^{\circ})$  to the farthest  $(5^{\circ})$ , we found a significant difference in performance (p<.001). In addition, when we compared the closest spacing (3°) to the no-flanker condition, that difference was also significant (p<.001), which replicates our result in Experiment 1. Critically, subject performance in the no-flanker condition was identical across the saccade and fixation conditions, indicating that the target face alone was easily recognizable. In addition, we observed that 43% of saccades to crowded faces landed within  $\pm 1.5^{\circ}$  of the center of the upright target face; an additional 51% landed within  $\pm 1.5$ -3°. This is nearly identical to saccades to unflanked faces where we observed 45% of saccades landed within  $\pm 1.5^{\circ}$  of the center of the upright target face; an additional 47% landed within  $\pm 1.5$ -3°.

#### Discussion

#### **Experiment 1**

This experiment provides the first evidence that executing a saccade to a crowded face improves recognition of that face. Notably, the improvement in identification performance that we report does not arise from direct foveation of the target face. The target face and its flankers were removed from the screen as soon as the subject's eye deviated from fixation by more than 0.5°, which prevented direct foveation of the face and, incidentally, resulted in subjects having considerably less time to acquire information about the face than in the fixation condition. In the saccade condition, the target face and its flankers were onscreen for an average of 70 ms, and yet subjects could still identify the face they were saccading to more accurately than when they maintained fixation on the cross and identified the face from peripheral information alone.

Although the results of Experiment 1 are intriguing, they do not necessarily indicate a reduction in crowding. The improvement in identification accuracy we observed in this experiment with saccades could be attributed to presaccadic attention alone, a result which would accord with previous findings (Deubel & Schneider, 1996; Kowler et al., 1995; Schneider & Deubel, 1995) although we should distinguish between presaccadic attention and simple covert attention in the absence of a saccade (see additional commentary in 4.2 and 4.3). To determine whether our procedure in Experiment 1 induced crowding and to probe the effect further, we performed a control experiment wherein we manipulated target-flanker spacing across both the saccade and fixation conditions (Experiment 2).

#### **Experiment 2**

In Experiment 2, we found that manipulating the target-flanker spacing as a diagnostic test for the presence of visual crowding resulted in decreased performance with smaller target-flanker spacings; this finding suggests that the stimuli used in Experiment 1 were made less identifiable by crowding and that we may have observed a partial release of crowding with saccades in that experiment. Our results in Experiment 2 show a pattern of decreasing accuracy with decreased target-flanker spacing, but crucially, we find a stable level of performance at larger target-flanker spacing across conditions, and we find an identical level of performance without flankers (Figure 2.3). Notably, the results in the saccade condition of Experiment 2 hint that there may be a partial saccade-mediated release of crowding with a saccade to a crowded target. We found a significant difference in performance between the saccade and fixation conditions at only one target-flanker spacing (3°) suggesting that a saccade to a crowded face does reduce the effects of crowding, but only in a limited range. We also found significant differences between the closest and furthest spacings in the fixation condition (2.5° and 4.5°) as well as between the closest spacing and the no-flanker condition, indicating that the inverted faces were effective in inducing crowding in our experiments. Given that performance in the noflanker condition was identical across the saccade and fixation conditions, our findings do not support a purely attentional explanation of the primary effect, as we would expect to see an increase in overall performance with saccades, even with minimal or

nonexistent crowding, which does not appear in our data. In addition, the significant difference in performance between the saccade and fixation conditions with 3° target-flanker spacing allows us to distinguish the effects of covert attention, as observed in the fixation condition, from those of presaccadic attention and other presaccadic processes. If the two forms of covert attention were identical, we would expect identical performance at all target-flanker spacings, as well as in the absence of flankers.

#### **General Discussion**

Our results suggest that not only does a saccade to a crowded face improve identification performance of isolated target objects, but that presaccadic attention and/or potentially other presaccadic processes triggered by the corollary discharge, as discussed by (Harrison, Mattingley, & Remington, 2013a), can provide sufficiently detailed information to discriminate between two faces. If discriminating two emotional faces requires a configural or holistic process (as suggested by (Maurer et al., 2002), then our results suggest that configural or holistic info can get through crowding in virtue of presaccadic processes including, perhaps, presaccadic attention. Our results extend prior work that was conducted with simple letter stimuli (Deubel & Schneider, 1996; Kowler & Blaser, 1995; Schneider & Deubel, 1995), rather than the faces we used. Presaccadic attention has been previously shown to be crucial for accurate saccade planning (Kowler & Blaser, 1995) and to improve identification performance with letter stimuli (Deubel & Schneider, 1996), but our work is the first to show that high-level, identifying information, such as is used to discriminate two faces, might be acquired prior to the saccade.

We believe that our results are consistent with those of Harrison and colleagues (Harrison, Mattingley, & Remington, 2013a), who found similar results with Gabor stimuli. Particularly, Harrison and colleagues found that saccadic amelioration of crowding was limited to close target-flanker spacings, and that saccades did not improve identification of weakly crowded targets. Our results suggest that crowded information about features or high-level objects in the visual periphery is, under some circumstances, available to the perceptual system when a saccade is prepared. This information may play a role in presaccadic identification of a saccade target. Acquiring detailed identifying information about a saccade target prior to the saccade itself may have significant implications for our understanding of visual stability (Melcher & Colby, 2008), and, more simply, for how we are able to readily identify objects in the world around us, particularly across eye movements.

The idea that saccade planning, including presaccadic attention, may capture more complex information is not without some precedent (Remington, 1980). For that matter, other work on visual crowding has shown that information about crowded stimuli is blocked from conscious access, but not discarded entirely (Dakin, 2001; Fischer & Whitney, 2011; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001). Our results suggest that presaccadic attention can make this information accessible in a way that partially bypasses or releases visual crowding. The information that is rendered inaccessible by visual crowding does not seem to be usefully retrievable in the absence of a saccade. Certainly, covert visual attention as used by our subjects in the fixation condition, is insufficient to relieve crowding to the degree we observe with saccades. However, presaccadic attention is not the only mechanism which might drive our effect. Other presaccadic processes, such as perisaccadic unmasking, as suggested by van Koningsbruggen and Buonocore (van Koningsbruggen & Buonocore, 2013) in their discussion of the work of Harrison and colleagues (Harrison, Mattingley, & Remington, 2013a), may be involved as well. We do not believe that perisaccadic unmasking can explain our effect; unlike the stimuli of Harrison and colleagues, our stimuli were not

masked, and were merely removed upon saccade onset. While we believe that our results suggest a distinction might be made between covert attention, as used in the fixation condition, and presaccadic attention, it is entirely possible that our effect is caused by other means, such as the nonattentional presaccadic enhancement suggested by Harrison and colleagues (Harrison, Mattingley, & Remington, 2013a).

In summary, these experiments demonstrate a presaccadic enhancement of crowded face recognition, and our results may represent an expansion of the known capabilities of presaccadic attention. Our results demonstrate that presaccadic attention or other presaccadic processes capture sufficient information to discriminate two crowded emotional faces, a task which requires considerably more information than discriminating the orientation of a letter or a Gabor patch. Presaccadic attention may be a privileged form of attention, since simple covert attention as used by our subjects in the fixation condition does not result in the same level of performance, and it appears to be crucial not only for saccade targeting (Kowler & Blaser, 1995), but potentially for identifying the target of the saccade before the eyes are in motion. Saccades have long been known to facilitate perception, but the act of making a saccade seems to enhance perception by acquiring information which is otherwise inaccessible prior to the saccade.

#### Chapter 3 : Saccadic Remapping of Object-Selective Information

Visual input across eye movements can be thought of as brief islands of stability in a storm of motion, yet we perceive the world as essentially stable. How we are able to do so has been a longstanding question for vision science (Helmholtz, 1898, trans. 1962). In recent years, saccadic remapping, where some visual neurons respond just before a saccade is about to bring a stimulus into their receptive field, has been suggested as a mechanism by which visual stability across saccades could be achieved. Early neurophysiological descriptions of remapping suggested that cells in lateral interparietal cortex shifted the locations of their receptive fields prior to a saccade (Duhamel, Colby, & Goldberg, 1992), and further work has suggested the corollary discharge (from the saccade planning process) may be key to this receptive field shift (Nakamura & Colby, 2002; Parks & Corballis, 2008; Sommer & Wurtz, 2008; Wurtz, 2008). Recently, it has been suggested that remapping may be a feature-selective process (Subramanian & Colby, 2014).

These neurophysiological findings have spurred psychophysical studies of saccadic remapping in the complete absence of direct recordings from neurons (Burr & Morrone, 2011; Cavanagh, Hunt, Afraz, & Rolfs, 2010; Melcher, 2005; 2007; Melcher & Morrone, 2003; Rolfs, Jonikaitis, Deubel, & Cavanagh, 2010). In these studies of perception, the term saccadic remapping usually refers to the presaccadic acquisition of information from the saccade target, with no necessary reference to a particular underlying physiological mechanism. We adopt a similar functional definition; we use the term saccadic remapping in this paper to refer to the perceptual consequences of presaccadic information, completely separate from its neurophysiological foundations.

Psychophysical studies of what is commonly called saccadic remapping can be thought of as suggesting one of two accounts of how remapping might interact with representations of the visual world in the brain. A retinotopic account holds that saccadic remapping is, in fact, the remapping of attention pointers for objects in the world immediately prior to a saccade within a retinotopic map, rather than the remapping of the representations themselves (Born, Ansorge, & Kerzel, 2012; Cavanagh et al., 2010; Collins, Rolfs, Deubel, & Cavanagh, 2009; Rolfs et al., 2010). Another view is that there are spatiotopic representations in visual cortex, representing objects in world centered coordinates independent of their current retinal position, and the representation of the object itself, rather than a pointer to it, is remapped (Burr & Morrone, 2011; Hall & Colby, 2011; Melcher, 2005; 2007; 2009; Melcher & Morrone, 2003).

However, previous research has focused more on how remapping might occur, relative to these representations of visual space in the brain, rather than what information might be presaccadically acquired. Saccadic remapping has been suggested to acquire feature-level information (Melcher, 2007; Subramanian & Colby, 2014), which would allow for only individual features to be remapped prior to the saccade, or coherent object representations could be remapped, which would facilitate postsaccadic comparisons and continuity of perception across saccades. Melcher, in an experiment on transsaccadic integration, had subjects foveate a male face for 5 sec, make a saccade away from the

face and, after an 800 ms delay, view a test (a morph between male and female) face for 250 ms, which they perceived as more female as a result of the presaccadic foveation (Melcher, 2005). This result suggests that object-specific information can be integrated across multiple viewpoints and that this information can influence postsaccadic perception. However, Melcher's 2005 work does not speak to what information might be acquired in saccadic remapping, since saccadic remapping only occurs very close to the time of the saccade (Harrison, Retell, Remington, & Mattingley, 2013b; Nakamura & Colby, 2002; Parks & Corballis, 2008; Rolfs et al., 2010; Sommer & Wurtz, 2008; Wurtz, 2008), and the adaptor was presented foveally prior to the saccade. More recent work suggests that feature-level information may be remapped (Melcher, 2007), but critically, these experiments relied on an initial foveal adaptor as well as a peripheral test after the saccade. Simply put, there has been no direct test of what information is remapped prior to a saccade using complex stimuli in the periphery as the saccade target. Related research has shown that saccades can reduce crowding with a variety of stimuli, including letters (Harrison, Mattingley, & Remington, 2013a) and even stimuli as complex as faces (B. A. Wolfe & Whitney, 2014). While this research suggests that detailed information is captured around the time of a saccade, no research to date has examined if detailed information can be acquired from a peripheral target, much less if saccadic remapping is an object-selective process.

Based on these studies, saccadic remapping might be an object-selective process; if so, there should be perceptual consequences of object-selective remapping with sufficiently complex stimuli. To approach this question, we used faces, which are natural, commonly encountered, and involve high-level object processing. We tested for a saccade-contingent sequential aftereffect in face perception, such as a negative face aftereffect or a sequential shape contrast effect. In addition, we tested whether this effect required holistic processing of faces, or if the sequential effect was based on a feature or features within our face stimuli. If saccadic remapping is an object-selective process, we might expect that perception of objects viewed postsaccadically would be modulated by information acquired prior to the saccade in an object-specific manner.

#### Methods

#### Display Setup

All experiments were performed using Matlab 2010a (Mathworks; Natick, MA), the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) and the Eyelink Toolbox (Cornelissen et al., 2002) on a Mac Mini (Apple; Cupertino, CA). Stimuli were displayed on a 47 cm Samsung cathode ray tube at a distance of 57 cm from the subject. For all experiments, monitor refresh was 100 Hz at a resolution of 1024 x 768.

#### Subjects

A total of 6 subjects participated in the experiments; the same 5 subjects participated in Experiments 1b-d and 2a-b. All subjects in all experiments had normal or corrected to normal vision. All subjects with the exception of one of the authors [BW] were naïve to the purposes of the study and provided written informed consent as required by the IRB at the University of California, Berkeley in accordance with the Declaration of Helsinki. All subjects were trained psychophysical observers and were accustomed to eyetracking experiments prior to data collection.

#### Stimuli

Stimuli were morphed emotional faces between the emotional states of happy, sad and angry, as originally used by Yamanashi Leib and colleagues (Yamanashi Leib et al., 2012). The morphs were generated by starting with two images of the same individual expressing either happy or sad emotional expressions, selected from the Ekman gallery (Ekman & Friesen, 1976). We then linearly morphed the faces to produce 48 morphs between each pair of basic emotions (i.e., 48 morphs between happy and sad). Morphs were created using Morph 2.5 (Gryphon Software, San Diego, CA), morphed between happy (morph 0) and sad (morph 49) with 48 intervening computer-generated morphs between them. Morph number 25 should be thought of as a neutral center-point within the morph space. In all experiments, the face stimuli, regardless of where they were presented on the display, subtended 4.26° high by 3.36° wide.



Figure 3.1: Stimulus sequence, Experiment 1. a: Experiment 1a, peripheral presentation of a happy or sad adaptor, followed by a saccade to the peripheral adaptor; judgment of morphed face after saccade. b: Experiment 1b, Foveal presentation of an emotional adaptor; identical in all respects to Experiment 1a. c: Experiment 1c, identical to Experiment 1a, with inverted stimuli throughout. d: Experiment 1d, identical to Experiment 1a, except that no saccade was permitted, and the test face was presented at the point of fixation for foveal assessment



Figure 3.2: Stimulus sequence, Experiments 2a, 2b. a: Experiments 2a and 2b were identical to Experiment 1, save for the addition of the second adaptor on each trial, as shown. There was always one happy and one sad adaptor presented during each trial and the hemifield they were respectively presented in varied randomly on a trial by trial basis. b: Experiment 2b varied the location of the saccade cue (red dot, on or near the cued target face), rather than locking it to the center of the adaptor, as in Experiment 1a-c and 2a.

#### Trial Sequence

Subjects' eye movements were recorded during all experiments, and they were not permitted to move their eyes from the fixation point (0.25 deg visual angle) except when instructed. All stimuli were presented gaze-contingently with a 0.5° threshold; if the point of gaze deviated by more than 0.5°, the stimuli were removed from the screen until gaze returned to the fixation point for a minimum of 500ms. Trials in all experiments began with an initial 1400 ms fixation period. In all experiments, the fixation dot was presented at the center of the screen and its Y-position varied randomly on a trial-by-trial basis within a 10 deg range from the center of the display.

The design of Experiment 1a (Figure 3.1a) was as follows: each trial started with a fixation period (1400 ms), followed by a peripheral adaptor (either a happy or sad face — morph 0 or 49, 15° to the left or right of fixation; Figure 3.1a) for 250 ms, and then a saccade cue (a 0.25° red dot 15° to the left or right of fixation, centered on the adaptor face). The adaptor face remained on the display while the saccade cue was presented. The adaptor and cue were both removed when the subject initiated the saccade, defined by the point of gaze deviating from fixation by more than  $0.5^{\circ}$ . Once the saccade was initiated, there was a 100 ms inter-stimulus interval, measured from the time of initial deviation from fixation (63 ms mean saccade duration for a 15° saccade), after which the morphed test face was presented. The test face, centered at the saccade cue location, was one of six possible faces, ranging from 15 to 35, centered at Morph 25 (between the happy and sad extremes); it was presented foveally after the saccade was complete. The test face duration was 100 ms, after which subjects were asked to judge whether they had perceived the foveal test face as happy or sad (2-alternative forced-choice task [2AFC]). Five subjects [four female; one author (BW); mean age 26.2] each performed one run of 360 trials.

In Experiment 1b (Figure 3.1b), a foveal adaptation control for Experiment 1a, the adaptor was placed at the initial fixation location, rather than in the periphery; all timing was identical to Experiment 1a. Subjects were instructed to make a saccade to the 0.25° saccade cue in the periphery. Subjects judged a test face identical to that used in the first experiment after making the saccade. Five subjects [four female; one author (BW), mean age 27.4] each performed one run of 360 trials.

Experiment 1c (Figure 3.1c) was an inversion control for Experiment 1a, and was identical in all respects to Experiment 1a, except that all face stimuli (adaptors and test faces) were rotated 180° from their orientations in Experiment 1a. Five subjects [four female; one author (BW); mean age 27.4] each performed one run of 360 trials.

Experiment 1d (Figure 3.1d) was a covert attention control, identical to Experiment 1a, except that subjects were not permitted to make a saccade to the adaptor, and the test face was presented at fixation to allow for foveal examination of the test stimulus, as in the saccade experiments. Five subjects [four female; one author (BW), mean age 27.4] each performed one run of 360 trials. Experiment 2a (Figure 3.2a) was an extension of the paradigm from Experiment 1; the only change was that two adaptor faces (one happy, one sad) were now presented on each trial (each was randomly assigned 15° to the left or right of fixation; Figure 3.2a). Subjects were presented with two synchronously presented saccade cues: a 0.25° green dot placed 1° to the left or right of fixation as a hemifield cue, and a 0.25° saccade cue (red), identical to that used in Experiment 1a, centered on the cued adaptor. The cues were always consistent with each other. Five subjects [four female; one author (BW), mean age 27.4] each performed one run of 360 trials.

Experiment 2b (Figure 3.2b) was a modified version of Experiment 2a; rather than having the saccade cue centered on one of the two adaptor faces (as described in Experiment 1a and 2a), the saccade cue could be placed anywhere within a 4° box centered on the adaptor face (Figure 3.2b). The test face was subsequently presented at the same onscreen location as the adaptor face following the saccade. This manipulation was to increase variance in saccade landing positions to better examine the spatial specificity of saccadic remapping. Five subjects [four female; one author (BW), mean age 27.4] each performed one run of 360 trials.

#### Eyetracking

Eyetracking was performed using an Eyelink 1000 (SR Research; Mississauga, Ontario, Canada) with a level desktop camera; data was recorded monocularly (right eye for all subjects) at 1000 Hz and saccade analysis was performed offline using the Eyelink parser. A saccade was defined as the first time point at which the velocity exceeded 30°/s and the acceleration exceeded 8000°/s<sup>2</sup>. In addition, a spatial displacement threshold was used to delay the start of each registered saccade until the eye had moved at least 0.15°. All subjects were stabilized on a chinrest during all experiments. Subjects were calibrated using a standard 9-point grid.

Eye movements were recorded at all times during all experiments; all stimuli were presented gaze-contingently. In experiments where a saccade was required (Experiments 1a-c, 2a-b), subjects were only permitted to make a saccade once the saccade cue was presented. Raw eye position was monitored continually throughout all experiments. Any deviation from fixation (greater than 0.5°) when the saccade cue was not present restarted the trial. If gaze deviated during an adaptation period, the adaptor(s) were removed from the screen until the subject foveated the fixation dot for a minimum of 500 ms. As a result, in experiments where a saccade was required, the subjects were unable to foveate the adaptor, since adaptor presentation was yoked to subjects' eye position. In Experiment 1d, where no saccades were made during the trial, any deviation greater than 0.5° resulted in the trial being discarded during analysis. Note that gaze-contingent stimulus control did not require saccades to be parsed; online monitoring of eye position was performed exclusively on the raw position samples from the eyetracker.

#### Analysis

All eyetracking and behavioral data were analyzed offline using custom Matlab scripts and S-R Research's EDFMEX data importation tool. The landing location of the first large saccade was calculated and used to filter the behavioral trials. In Experiments

1a-c and 2a, trials with saccades that landed greater than  $2^{\circ}$  from the center of the face were discarded (on average, 9% of trials in experiments requiring a saccade). Saccade errors greater than  $2^{\circ}$  were analyzed in Experiment 2b. Saccade landings in Experiment 2b were converted into a single linear vector of deviation from the center of the target face (face-centered), rather than screen-centered landing location, and were filtered using a overlapping progressive exclusion procedure with  $0.5^{\circ}$  radius steps from the center of the face to better assess the spatial tuning of saccadic remapping. Accordingly, trials in Experiment 2b where the deviation vector was  $0-1^{\circ}$  were analyzed separately than trials where the deviation vector was  $0.5^{\circ}$ . (in steps of  $0.5^{\circ}$ ) to a maximum deviation bin of 2-3°, incorporating 96% of all trials in the experiment. This served to smooth the data given the greater imprecision in saccade landing resulting from the randomly-shifted saccade cue.

The 2AFC psychophysical data were sorted by adaptor condition and the data were fit with the logistic function  $(x) = \frac{1}{1+e^{-\alpha(x-\beta)}}$ , where parameter  $\alpha$  is the slope and parameter  $\beta$  is the point of subjective equality (PSE). This generated two logistic functions per experiment per subject; the difference between the PSEs of these two functions is the size of the face aftereffect (see inset, Figure 3.3). While the slope was allowed to vary, differences in slope between functions were not examined, as the face aftereffect is one of appearance, rather than a change in sensitivity. Error bars in Experiments 1a-d and 2a are bootstrapped 95% confidence intervals, and significance was tested (2-tailed p-values) using a bootstrapping procedure on individual subjects (Efron & Tibshirani, 1993); data were averaged after individual bootstrapping to assess group performance. Error bars for Experiment 2b are one standard deviation, using the same procedure.

#### **Experiment 1: Saccadic Remapping of Faces**

To test if saccadic remapping is an object-selective process, which is to say, if more than an isolated feature was remapped prior to the saccade, we performed four initial experiments with a single adaptor. In Experiment 1a, subjects were presented with a peripheral adaptor that was never foveated, which was gaze-contingently removed from the screen on saccade onset. If we observed any face aftereffect in this experiment, which would require the complete representation of the face to be remapped, it would indicate that object-selective information is acquired via saccadic remapping. Experiment 1b, with a foveal adaptor prior to a saccade to the peripheral saccade cue, determined the maximum aftereffect possible with foveal adaptation rather than saccadic remapping of a nonfoveated adaptor. Experiment 1c was a replication of the saccadic remapping experiment (1a) with inverted stimuli to test for holistic, therefore object-selective, (Fischer & Whitney, 2011; Maurer et al., 2002) versus featural (Valentine, 1988; Yin, 1969) processing of remapped faces. Experiment 1d was a covert attention control for the saccadic remapping experiment, where no saccade was permitted to the peripheral adaptor. In all experiments, subjects judged a foveally-presented test face at the end of the trial.

#### Results

Experiment 1a, with a never-foveated adaptor, reveals an average aftereffect of 2.21 morph units (2-tailed bootstrap t-test; p<.001; Figure 3.3). With the foveal adaptor in Experiment 1b, the average aftereffect was 5.32 morph units (2-tailed bootstrap t-test; p<.001). With the inverted stimuli in Experiment 1c, the average aftereffect was 0.853 morph units (2-tailed bootstrap t-test; p=.008). Experiment 1d, the covert attention control, showed a nonsignificant average aftereffect of -0.50 morph units (p>.05).



**Figure 3.3**: Results, group effect size for Experiment 1a-d; Experiment 1a, remapping of a peripheral target, shows a significant face aftereffect (p<.001); Experiment 1b, with a foveal adaptor, shows significant adaptation when a face is briefly foveated (p<.001). Experiment 1c, with inverted stimuli, shows a significant (p=.008) but reduced aftereffect compared to Experiment 1a (p=0.035). Experiment 1d, covert attention control for Experiment 1a, shows no significant aftereffect (p>.05). Error bars are bootstrapped 95% confidence intervals. Inset shows two representative psychometric functions for one subject in Experiment 1a; the difference between the two functions, independent of any differences in slope, represented by the arrow, is the size of the effect.

#### **Experiment 1 Discussion**

In Experiment 1a, the adaptor face was presented in the periphery, but it was gaze-contingently removed from the screen on saccade onset. Thus, the adaptor was never foveated. Only the test face was ever foveated. Despite presenting the adaptor and test faces to different retinal locations, there was a strong negative face aftereffect. Thus, the aftereffect we observe in this experiment must have been induced by information acquired during the saccade planning process. Our result in this experiment is the first to show that perception of a face can be changed by remapped information, suggesting that remapping is object-selective.

Experiments 1b-d are control experiments for Experiment 1a, to rule out alternative explanations of the effect. Our results in Experiment 1b are in accord with previous findings demonstrating induction of a face aftereffect with brief foveal adaptation (Leopold, Rhodes, Müller, & Jeffery, 2005; Rhodes, Jeffery, Clifford, & Leopold, 2007). Experiment 1c, with inverted stimuli, addressed whether saccadic remapping is an object-selective or feature-selective process. It was possible that subjects in Experiment 1a acquired a salient feature from the stimuli that characterized the emotion of the face (e.g., the visibility of teeth in a smiling face versus their absence in a sad face). If this had been the case, we would have expected to find a significant aftereffect in Experiment 1c similar to what we found in Experiment 1a. We find a significantly reduced but significant aftereffect with inverted stimuli as expected from the preexisting literature (McKelvie, 1995; Rhodes, Evangelista, & Jeffery, 2009; Rutherford, Chattha, & Krysko, 2008; Yin, 1969). Experiment 1d, our covert attention control, showed that a saccade is necessary and covert attention alone is insufficient to generate the negative face aftereffect in Experiment 1a. The lack of an effect in Experiment 1d also suggests that our results in Experiment 1a cannot be accounted for by a global face aftereffect, as would be suggested by Afraz and Cavanagh (2008). Our results in Experiment 1 demonstrate that saccadic remapping is an object-selective process that cannot be accounted for by covert attention or feature remapping alone.

## **Experiment 2: Saccadic Remapping with Multiple Targets Experiment 2a: Target Specificity of Saccadic Remapping**

Some negative aftereffects spread globally and are therefore spatially non-specific (Afraz & Cavanagh, 2008; McKone, Jeffery, Boeing, & Clifford, 2014). Conversely, remapping is believed to be a spatially specific process (Duhamel, Colby, & Goldberg, 1992; Hall & Colby, 2011; Rolfs et al., 2010; Wurtz et al., 2011). If the effect we find in Experiment 1a is due to remapping, rather than a global effect, we should expect spatial specificity of the aftereffect from the remapped location. To address the spatial specificity of the face aftereffect, we presented two adaptor faces, one in each visual field with opposite emotions (see methods). If, in fact, our results could be accounted for by a globally-induced negative aftereffect, as suggested by Afraz and Cavanagh, we would expect to see no effect at all in this experiment, as the effects of each adaptor would cancel.



**Figure 3.4**: Experiment 2 Results. **a**: Group mean results for Experiment 2a, with two adaptors on screen in each trial, and for comparison, Experiment 1a (from Figure 3.3). Both results are respectively significant, p<.001, and there is no significant difference between them (p>.05). **b**: Results of the progressive overlapping exclusion analysis for Experiment 2b. We find a strong presaccadically-induced face aftereffect only when the deviation between saccade landing and the center of the target face is between within 0.5° and 1.5° of the center of the face (asterisk indicates p<.001; Bonferroni-corrected  $\alpha = .01$ ; note that the 0-1° error bin is trending at p = .015). Given the size of the stimulus (3.6° wide by 4.3° high), this indicates that the saccade must land where the adaptor face had been. When saccade error is higher, we observe no significant aftereffect, p>.05 (in the 1°-2°, 1.5°-2.5° and 2°-3° saccade error bins). Error bars for this analysis only are bootstrapped 1SD.

#### Results

We find an aftereffect in Experiment 2a of 2.16 morph units with a (2-tailed bootstrap t-test; p<.001) (Figure 3.3). Additionally, we found no significant difference between the size of the effect in this experiment and in Experiment 1a (2-tailed bootstrap t-test, p>.05).

## **Experiment 2b: Spatial Tuning of Saccadic Remapping**

Experiment 2a suggests some degree of spatial tuning, as the effect was undiminished by the presence of the adaptor with the opposite expression in the other hemifield. To more precisely characterize the spatial tuning of object-selective remapping, we varied the location of the saccade cue in Experiment 2b, rather than simply centering it on the adaptor face, as we had done in previous experiments (as illustrated in Figure 3.1a-d and Figure 3.2a). The saccade cue, to which subjects targeted their saccade, was presented at a random location in a  $\pm 2^{\circ}$  box (4° on a side) centered on the adaptor (illustrated in Figure 3.2b). As a result, saccade landings in Experiment 2b were more widely distributed, relative to the adaptor face, than in Experiments 1a-c and 2a. This allowed us to analyze the spatial tuning of the effect. Note that the test face was presented at exactly the same location as the adaptor face; subjects' saccade landing locations varied as to their proximity to the test face.

For each trial, we calculated the vector between the subject's saccade landing location and the center of the adaptor face. This allowed us to perform a progressive, overlapping exclusion analysis (shown in Figure 3.4b and described in Methods); trials were binned by spatial accuracy relative to the location of the adaptor face in overlapping one degree bins (i.e., trials where the saccade landed 0-1° from the center of the face, 0.5-1.5° from the center of the face, etc.). Thus, trials fell into multiple bins, effectively smoothing the data. While such an analysis is possible in our previous experiments, the high precision of saccade landings in those experiments limited the number of high-error trials available for this type of analysis.

#### Results

On trials where the saccade landed within 1° of the center of the adaptor location (Figure 3.4b), the aftereffect was 2.06 morph units (p=.015; trending but not significant at a Bonferroni-corrected  $\alpha = .01$ ), consistent with the previous experiments. Trials where the saccade landed between 0.5-1.5° from the center of the adaptor location showed a significant aftereffect of 2.16 morph units (p<.001; significant at a Bonferroni-corrected  $\alpha = .01$ ) When saccade landings were further from the face, there was a decrease in the strength of the aftereffect; landing deviations greater than approximately 1.5 deg resulted in little measurable aftereffect. All individual subjects in this experiment displayed a decrease in aftereffect magnitude with greater saccadic imprecision. In addition, we performed a similar analysis in the absence of overlapping bins (i.e., bins from 0-1°, 1-2°) and found a similar pattern, with no significant aftereffect when saccades landed more than 1° from the center of the face.

## **Experiment 2 Discussion**

Experiments 2a and 2b revealed a saccade-contingent negative aftereffect with the presence of two adaptors on each trial, which rules out the possibility that the effect we observed in Experiment 1a was the result of global spatially-nonspecific adaptation or other spatially non-specific decision biases. If our effect was caused by a spatially non-specific process (something other than remapping), we would expect the two opposing adaptor faces to reduce or eliminate the negative aftereffect. Experiment 2b shows that not only are saccades required for the remapped negative face aftereffect, but the accuracy of the saccade targeting modulates the aftereffect (Figure 3.4).

#### **General Discussion**

In this series of experiments, we have demonstrated that, prior to a saccade, a complete representation of a target object is acquired by the visual system, sufficient to induce a negative face aftereffect, rather than individual features of our stimuli, which would not have resulted in the effect we observed. Our findings reinforce the idea that saccadic remapping, in the sense we have used the term, may operate in a spatiotopic framework by remapping object representations (Burr & Morrone, 2011), rather than the remapping of attention pointers in a purely retinotopic framework (Rolfs et al., 2010), although both models could accommodate our results. However, our primary interest concerned what information was acquired from the saccade target prior to the saccade itself, and our results suggest that more information is acquired than was previously suspected. This remapped information from a saccade target, acquired in an object-selective manner, may facilitate the postsaccadic updating of our perceptual representation of the visual world.

Several control experiments rule out alternative explanations for our results. Experiment 1c showed that it is configural or holistic (i.e., object-level) information about faces that is remapped and produces a negative face aftereffect. Thus, it is unlikely that features (e.g., teeth) or a figural aftereffect (Rhodes et al., 2004) could explain our results. Experiment 1d demonstrated the necessity of saccades; simply attending to a peripherally presented face (as in Experiment 1d) does not allow the visual system to acquire sufficient information to change perception of a subsequent face. Further, the remapped face aftereffect was spatially specific (Experiments 2a and 2b), ruling out decision biases, non-spatially specific processes, global aftereffects (Afraz & Cavanagh, 2008), or other non-remapping selective processes (Deubel & Schneider, 1996; Harrison, Mattingley, & Remington, 2013a; Kowler et al., 1995).

Our method for determining whether saccadic remapping is object-selective is based on how the perception of a foveated neutral face changes, depending on what, if any, information is acquired before it is presented. We have chosen to describe this as a negative aftereffect, building on the literature surrounding the face aftereffect (Webster, Kaping, Mizokami, & Duhamel, 2004), but it could also be characterized as an uprightface-specific figural aftereffect (Asch & Witkin, 1948) or shape contrast illusion (Suzuki & Cavanagh, 1998), particularly given the brevity of our presentation. However, there is little reason to expect an inversion effect for a figural aftereffect, as we observed in Experiment 1c. While there may be debate over what our effect may best be called, the fact remains that subjects' perception of the test face only changes when they have made a saccade to an emotional face that they never foveate, indicating that saccadic remapping, as assessed through psychophysical methods, is an object-selective process.

Related to this, recent work on the serial dependence in face perception (Liberman, Fischer, & Whitney, 2014) has shown a positive shift in how subsequent faces are perceived over a longer timescale. This suggests that the continual process of constructing a stable perception of the world can be influenced in multiple ways. Our negative aftereffect from object-selective remapping reflects information available to the visual system immediately prior to the time of the saccade, whereas serial dependence of faces is a longer-term process, on the order of seconds rather than milliseconds. It may be that transsaccadic comparisons exaggerate differences between remapped and foveated information, maximizing sensitivity to changes, whereas serial dependence subserves a stable representation of objects in the world.

Saccadic remapping of object-selective information, which is a spatially tuned process, may facilitate object perception across eye movements by acquiring sufficient information prior to a saccade to permit immediate comparison after the saccade target is fixated. Acquiring information at the level of objects around the time of a saccade may help explain object stability across saccades (Collins, Heed, & Röder, 2010; Rensink, 2002; Wurtz et al., 2011) and may also speak to the speed with which objects can be recognized (Greene & Oliva, 2009; Oliva & Torralba, 2007; Pajak & Nuthmann, 2013). Immediately prior to a saccade, the visual system acquires object-specific information, and may use this information after the saccade to update representations, simultaneously maximizing sensitivity to any potential object changes while facilitating stable perception of unchanging objects.

Chapter 4 : Foveal Input is not Required for Perception of Ensemble Facial Expression

Our visual world is composed of complex information that is continually changing from moment to moment. Any given scene contains a wealth of visual information—pebbles on a beach, leaves on a tree, faces in a crowded room—yet limitations on our attention and short-term memory prevent us from processing every detail (Luck & Vogel, 1997; Duncan: 1994it, Myczek & Simons, 2008). One way in which the visual system is able to efficiently process this information is by extracting summary statistics (e.g., the average) of a given stimulus feature across an array of objects through a process known as ensemble perception (for reviews, see (Alvarez, 2011; Fischer & Whitney, 2011; Haberman & Whitney, 2011; Haberman, Harp, & Whitney, 2009). A large body of evidence has shown that the visual system can rapidly extract the mean of stimulus features such as orientation (Ariely, 2001; Dakin & Watt, 1997; Parkes et al., 2001), size (Ariely, 2001; Carpenter, 1988; Chong & Treisman, 2003), and motion direction (Watamaniuk & Sekuler, 1992). In recent years, further research on the topic has demonstrated that observers can perceive the mean features from complex objects, such as crowd heading from point-light walkers (Sweeny, Haroz, & Whitney, 2013), emotions from sets of faces (Haberman et al., 2009; Haberman & Whitney, 2007; Ji, Chen, & Fu, 2013; 2014; Jung, Bulthoff, Thornton, Lee, & Armann, 2013; Yang, Yoon, Chong, & Oh, 2013), facial identity (de Fockert & Wolfenstein, 2009; Haberman & Whitney, 2007; Yamanashi Leib et al., 2012; 2014), crowd gaze direction ((Cornelissen et al., 2002; Sweeny & Whitney, 2014)), and auditory tone (Piazza, Sweeny, Wessel, Silver, & Whitney, 2013). However, it remains a debated question whether ensemble perception of high-level visual stimuli, such as faces, can be accomplished covertly or if it requires overt, sequential foveation of objects before an ensemble representation can be extracted.

Ensemble perception could result from a covert process in which coarse but sufficient information is gathered from the periphery to generate an ensemble percept (Fischer & Whitney, 2011; Haberman et al., 2009; Sherman, Evans, & Wolfe, 2012). Consistent with this, ensemble perception of simple features (e.g., size, orientation) has been shown with a range of brief stimulus durations (from 50 to 500 ms, c.f. (Ariely, 2001; Dakin & Watt, 1997; Parkes et al., 2001), providing some support for the covert account, as the stimulus durations are often shorter than the time required to plan a saccade (approximately 200 ms; (Carpenter, 1988). In addition, more recent evidence demonstrates ensemble perception of more complex features with brief stimulus durations. For instance, Sweeny and colleagues (Sweeny et al., 2013) demonstrated that observers can extract the average heading from groups of point-light walkers with durations as short as 200 ms, and Yang and colleagues (Yang et al., 2013) showed ensemble processing of emotional faces with a 100 ms stimulus duration.

Conversely, ensemble perception might rely on an overt process, where sequentially fixated stimuli are averaged. Indeed, recent experiments suggest a dominant, if not necessary, role for foveally presented faces when extracting ensemble expression or identity (Ji et al., 2013; 2014; Jung et al., 2013). The drawback of all of the studies above—whether they support overt or covert ensemble representations—is that they are indirect tests. The most direct test for the necessity of foveal input when extracting ensemble crowd expression is to simply block the fovea in a gaze contingent manner. We performed such a series of experiments in which subjects were asked to report the average emotion of a group of faces without foveal input. Using high-speed eyetracking and gaze-contingent stimulus control, we occluded the central 2.6° of the visual field, entirely blocking foveal input. Subjects performed an ensemble perception task in which they were asked to report the mean emotion of a group of faces by matching a test face to the previously seen group. We compared subjects' performance when the foveal occluder was present to a control condition in which the occluder was absent; if foveal information is not necessary, we would expect identical performance in the two conditions. In a second experiment, we utilized a subset design to measure how much face information observers are able to integrate from the display with and without foveal input.

## Experiment 1: Ensemble Perception of Facial Emotion with and without Gaze-Contingent Foveal Occlusion

To test the role of foveal input in creating an ensemble percept, we performed an experiment where subjects were asked to determine the mean emotion of an array of 24 emotional faces (Figure 4.1). In one condition, subjects were able to freely view the stimulus array without interference, during which their eye movements were recorded. In a second condition, using online gaze position data from the eye tracker, we occluded the foveal region of the visual field. The occluder (Figure 4.1c) consisted of a white patch with a flattened Gaussian luminance profile (to blend seamlessly into the background) resulting in a circular area (2.6 degrees in diameter) of full occlusion. Subjects matched the mean emotion of the presented faces in both conditions by scrolling through the entire face space (Figure 4.1b; 147 total faces) using the method of adjustment, and clicking on the matching face.

#### Experiment 1 Methods

#### Subjects

Six subjects (including two authors, 4 female; mean age 26.7) participated in this experiment. All subjects reported normal or corrected-to-normal vision. Subjects provided written informed consent as required by the Institutional Review Board at the University of California, Berkeley in accordance with the Declaration of Helsinki. Aside from the two authors who participated (AK and KW), subjects were naïve to the purpose of the experiment.

#### Display Setup

Stimuli were presented on a 43 cm Samsung SyncMaster 997DF cathode ray tube with a monitor refresh rate of 75 Hz and a resolution of  $1024 \times 768$ . Subjects were seated in a dark booth at a viewing distance of 57 cm from the monitor and head movement was limited with a chinrest. At this distance, 30 pixels subtended approximately 1° of visual angle. The experiment was run on a Mac Mini (Apple; Cupertino, CA) and written using Matlab 2010a (Mathworks; Natick, MA) used in conjunction with the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) and the Eyelink Toolbox (Cornelissen et al., 2002).



**Figure 4.1:** (A) Stimulus space used in all experiments (3 emotions; happy, sad and angry; 47 morphs between each emotion). (B) Illustration of foveal occluder (Gaussian blob; 1.125° SD) relative to size of stimulus face. (C) Stimulus sequence for Experiment 1. Subjects began each trial by fixating a cross in the center of the screen for 700-1500 ms, after which time the stimulus array (24 emotional faces, see Methods) was presented for 1500 ms of free viewing. The fovea was either occluded (blue) or non-occluded (red). After 1500 ms, the stimulus was removed and subjects were able to adjust a face presented onscreen by moving the mouse through the entire 147-face space.

#### Stimuli

Stimuli were morphed faces between the emotional states of happy, sad and angry, as used by Yamanashi Leib and colleagues (Yamanashi Leib et al., 2012). The morphs were generated by starting with three images of the same individual expressing happy, sad, or angry emotional expressions, selected from the Ekman gallery (Ekman & Friesen, 1976). We then linearly morphed the faces to produce 48 morphs between each pair of basic emotions (i.e., 48 morphs between happy and sad, 48 morphs between sad and angry, and 48 morphs between angry and happy), for a total of 147 faces (Figure 4.1a, 144 morphs plus 3 original images). Note that the set forms an approximately circular continuum where there is no beginning or ending face. Morphs were created using Morph 2.5 (Gryphon Software, San Diego, CA). The grayscale images of the faces were ovals 2.93° wide by 3.65° high, and cropped so that hair and other background features were not visible. The mean luminance of the faces was 57.7 cd/m<sup>2</sup> and their mean contrast was 94%. Faces were presented on a white (141.5 cd/m<sup>2</sup>) background.

On each trial, the set of 24 presented faces was arranged in two concentric rings (Figure 4.1c), consisting of an inner ring of 9 faces and an outer ring of 15 faces. The inner and outer rings had radii of  $6.75^{\circ}$  and  $10.5^{\circ}$ , respectively, from the center of the display. To add random variation to the positions of the faces on each trial, the center of each face was randomly jittered around a set of evenly spaced angular locations within each ring. On each trial, faces in the inner ring were jittered by up to  $\pm 4.74^{\circ}$  of rotation angle; faces in the outer ring were randomly jittered by up to  $\pm 2.4^{\circ}$  of rotation angle, maintaining their distances from the center of the display. This amount of position jitter prevented any overlap or occlusion of the faces.

On each trial, 24 face morphs were selected from a Gaussian probability distribution. The center of the distribution (i.e., the average face) was selected with uniform probability from the full set of 147 morphs, and the standard deviation of the distribution was always 18 morphs. In addition, to assess any potential effects of display arrangement on subjects' performance, we performed an additional stimulus manipulation by organizing the faces by their emotional state, similar to the work of Sherman and colleagues (Sherman et al., 2012), who found organization-based facilitation of ensemble perception of orientation. On the organized trials, the face closest to the mean was assigned a random location out of 24 possible positions with equal probability. The remaining faces were sorted by absolute morph distance from the average face and assigned to the remaining 23 slots based on angular distance from the slot containing the face closest to the mean, with morph separation increasing with angular distance. On the random trials, the selected morphs were selected randomly assigned to each slot. As we found no difference in mean error between the organized (13.28 morph units) and random (13.48 morph units) conditions at the group level (p = 0.89) in Experiment 1, all data were averaged across the organized and random conditions in all subsequent analyses.

#### Trial Sequence

On each trial, subjects fixated a  $0.23^{\circ}$  black cross  $(1.9 \text{ cd/m}^2)$  at the center of the screen for a random period, between 700 and 1500 ms, and were subsequently shown the

array of 24 emotional faces for 1500 ms. Subjects were allowed to freely move their eyes around the screen (Figure 4.1c). After the stimulus was removed from the screen, a 200 ms inter-stimulus interval (ISI) elapsed before subjects were shown a single face that they were instructed to adjust to match the mean emotion of the previously presented faces. Using a mouse, subjects were able to adjust the face on the response screen to any one of the 147 morphs. Once subjects had entered their response by adjusting the face to the perceived mean, and clicking the mouse to confirm their response, an 800 ms inter-trial interval (ITI) elapsed before the next trial. Subjects were given feedback on their performance. Responses within 20 morphs from the mean of the set resulted in a highpitched (652.9 Hz) tone, indicating an accurate response, and responses more than 20 morphs from the mean resulted in a low-pitched (157.1 Hz) tone, indicating an inaccurate response. With the exception of the presence of the gaze-contingent occluder at the fovea (Figure 4.1b), the procedure was identical across the occluded and non-occluded conditions. Subjects performed the task in 6 blocks of 80 trials each-3 blocks in the non-occluded condition and 3 blocks in the occluded condition, for a total of 240 trials per condition. To avoid training effects, the sequence of conditions was randomized across subjects, with half the subjects running in the occluded condition first and the other half running in the non-occluded condition first.

#### Eyetracking

Subjects' eye movements were recorded throughout each run using an Eyelink 1000 (SR Research; Mississauga, Ontario, Canada) with a level desktop camera, recording the right eye at 1000 Hz. Subjects were calibrated using a standard 9-point grid (mean error <0.5°). For the fixation analysis (see Results), time points from the recording were parsed into fixations and saccades offline using the Eyelink parser. The beginning of a fixation interval was defined as the first time point where at which the velocity fell below  $30^{\circ}$ /s and the acceleration fell below  $8000^{\circ}$ /s<sup>2</sup>, and saccades were defined as time points in which velocity and acceleration exceeded their respective thresholds.

In the foveal occlusion condition, we used the raw gaze position data from the eyetracker to present a white occluder with a flattened Gaussian luminance profile according to the equation:

$$f(x,y) = \begin{cases} A, & \text{if } x_0 + z > x > x_0 - z \\ 2A\exp\left(\frac{-((x-x_0)^2 + (y-y_0)^2)}{2\sigma^2}\right), & \text{otherwise} \end{cases}$$

where x and y represent horizontal and vertical position, respectively,  $x_0$  and  $y_0$  represent gaze position,  $\sigma$  represents the standard deviation, A represents the amplitude (corresponding to the maximum luminance of the patch), and z represents the location of the full-width at half-maximum (FWHM) or:

$$z = \sqrt{(2 \ln 2)\sigma^2 - (y - y_0)^2}$$

The minimum luminance of the occluder was identical to the background (141.5 cd/m<sup>2</sup>), allowing it to blend seamlessly in with the background, and the standard deviation was set to 1.125°, resulting in a fully occluded region approximately 2.6° in diameter. The dimensions of the occluder were determined based on both the dimensions of the stimuli and retinal anatomy. The central 2.6 degrees of the visual field has an area

approximately twice as large as the entire rod-free portion of the fovea (1.8° diameter; (Poliak, 1941). In addition, at the edge of the fully occluded-region (i.e., 1.3 degrees eccentricity), human cone density drops to approximately 22.8% of its maximum (Curcio, Sloan, Kalina, & Hendrickson, 1990). More importantly, an occluder of this size fully covers the features (eyes, nose, and mouth) of each face when fixated centrally (see Figure 4.1b for a scale comparison of the occluder and an example face). This way, subjects were unable to extract detailed features of the face images when fixating them directly.

#### Analysis

All data (behavioral and eyetracking) were analyzed offline using custom Matlab scripts and S-R Research's "edfmex" file import tool. For the behavioral responses, we calculated the absolute difference between the mean emotion of the presented faces and the subject's chosen match face on each trial and then calculated the mean across trials to get a measure of subjects' errors across the different conditions. We performed nonparametric bootstrap tests in order to compare subjects' performance between the occluded and non-occluded conditions (using the method of (Efron & Tibshirani, 1993). Bootstrapped estimates of mean response error were calculated by resampling each subject's data 1000 times with replacement separately for the occluded and non-occluded conditions. Bootstrapped estimates were then averaged across subjects to compare response errors between the occluded and non-occluded conditions. To compare observers' performance to chance (i.e., floor) performance, we calculated a null distribution of the expected errors generated by random guessing. For each of the 1000 permutations, we shuffled the mapping between the mean of the presented group and subjects' responses and recalculated the error. In other words, the error on each trial was calculated by comparing the mean of the presented group on one trial to the response on a different trial.



**Figure 4.2:** (A) Mean Absolute Response Error result for Experiment 1. In Experiment 1, we found no significant difference with six subjects (p=.874; bootstrapped, 2-tailed) between the mean of absolute response error in the foveal occlusion case and the non-occluded case. (B) The distribution of absolute response error in the occluded (upper) and non-occluded (lower) conditions for a single exemplar subject.

#### Experiment 1 Results

#### Response errors

To determine the effects of display configuration on subjects' performance, we compared the mean of the absolute errors between the random and organized display conditions. There was no difference (p = .22). In addition, we find no difference between the foveal occlusion (absolute mean error, 13.69 morph units) and the non-occluded (absolute mean error, 13.17 morph units) conditions (p = 0.874) (Figure 4.2a). In addition, there was no effect of block order; the difference in absolute mean errors between the foveal occlusion and non-occluded conditions was similar for subjects that performed the occluded condition first versus those that performed it second (p = 0.246). Importantly, the lack of difference between these two conditions is not because of chance or floor performance. Subjects were very sensitive to average expression of the crowd, and performance was significantly above the expected chance performance level of 36.75 morph units (mean of permuted distribution; permutation test, p < .001), replicating several previous studies (Haberman et al., 2009; Haberman & Whitney, 2007; 2009; Sweeny et al., 2013).

#### Eye tracking.

In addition to examining whether the presence of a foveal occluder affected performance, we compared fixation behavior between the occluded and non-occluded conditions. For the eye tracking analysis, we analyzed the fixation locations for each trial and determined whether they corresponded to a fixation on or off of one of the faces onscreen during that time, and then calculated the proportion of trials where subjects did not directly foveate a face. There were some slight differences (at a Bonferroni-corrected  $\alpha = 0.0125$ ) between the occluded and non-occluded conditions in the mean number of total fixations per trial (occluded = 5.31, non-occluded = 5.39, p = 0.006 (non-occluded > occluded) and fixation duration (occluded = 186.79 ms, non-occluded = 189.82 ms, p = 0.01 (non-occluded > occluded). In addition to these common saccade and fixation metrics, we also examined whether there were any differences in gaze position relative to the faces. In particular, we classified each time point recorded during the 1500 ms stimulus presentation by whether or not the subject's point of gaze overlapped with any of the 24 faces. We then calculated the proportion of the stimulus duration (out of the 1500 ms of the trial) during which subjects fixated the space in between the faces in the display. This duration was longer in the foveal occlusion condition versus the nonoccluded condition (occluded = 513.74 ms, non-occluded = 385.65 ms; p < 0.001), indicating differences in saccade behavior across the two conditions (Figure 4.2b). Similarly, we classified each parsed fixation event based on whether it overlapped with one of the 24 faces, and for each trial calculated the mean proportion of fixations that were not directly on a face. Consistent with the mean duration result, there was a greater proportion of fixations in the gaps between the presented faces in the occluded condition compared to the non-occluded condition, (occluded = 0.3361, non-occluded = 0.2392; p < 0.001).

#### **Experiment 1 Discussion**

Experiment 1 compared observers' accuracy in reporting the mean emotion of a set of faces, with and without a foveal occluder, to test whether foveal information is necessary for ensemble perception. In the occlusion condition, subjects were prevented from extracting detailed foveal information from any face, yet they were able to perform our ensemble perception task just as accurately as when no aspect of the stimulus was occluded. If ensemble perception relied on averaging foveal information from sequential fixations, we would have expected the occluder to have significantly impaired subjects' performance. Therefore, our results suggest that ensemble perception of facial emotion does not require foveal input, in contrast to previous reports (Ji et al., 2013; 2014; Jung et al., 2013).

However, we do find a significant difference in the amount of time subjects spend fixating faces directly between our two conditions. Given that subjects are not able to acquire detailed information about facial expression when fixating the faces in the occlusion condition, it is unsurprising that subjects opt to maximize the available information by fixating in the interstitial space between the faces. This is not to say that subjects exclusively fixated between faces in the occlusion condition; the majority of their fixations (66.39%) remained targeted at faces, rather than interstitial space. Despite this change in behavior or strategy, the results suggest that foveal detail is simply not required to process ensemble information.

#### **Experiment 2:** Do observers integrate information from multiple faces?

While our results in Experiment 1 suggest that a lack of foveal input does not necessarily impair observers' ability to perceive the mean emotion of an array of faces, the amount of information that observers use to compute this average in the occluded relative to the non-occluded condition remains an open question. One possibility is that observers extract a representation of the average emotion of the crowd by integrating information across the group. Another possibility is that subjects may be using as little as one face from the set to make their judgment. In order to test whether multiple faces contribute to subjects' ensemble judgments, we modified Experiment 1 to present a subset of the total faces and calculated subjects' errors relative to the entire set of 24 faces. If subjects are only using one face from the set of 24 to make their judgment, performance (when errors are calculated relative to the full set) should be the same when only one random face is visible compared to when all 24 faces are visible. If they are using a larger number (e.g., 8 faces) to make their judgment, performance with a random subset of 8 should be better than when only one face is visible. In other words, we expect that if subjects integrate information from multiple faces, performance should improve with an increasing number of faces presented. If subjects do not integrate multiple faces, we would expect that adding additional faces would not improve performance in the task.

#### Experiment 2 Methods

#### Subjects

Four subjects (2 authors; AK and KW; all female; mean age 25) who participated in Experiment 1 also participated in this experiment. Subjects provided written informed consent as required by the IRB at the University of California, Berkeley in accordance with the Declaration of Helsinki. Aside from the two authors who participated (AK and KW), subjects were naïve to the purpose of the experiment.



**Figure 4.3:** Experiment 2 Stimuli. On any given trial (A), subjects could be presented with 1, 2, 4, 8 or 12 faces (B) from a total set of 24 and asked to report the mean emotion of the set. The methods were otherwise similar to Experiment 1.

#### Stimuli and Procedure

The stimuli (Figure 4.3a) and procedure in Experiment 2 were identical to those of Experiment 1 aside from the elimination of the organized configuration and the addition of a subset design. First, given the lack of a display configuration (random vs. organized) effect in Experiment 1, only the random condition was used in Experiment 2. In other words, the faces were randomly assigned to the 24 "slots" on each trial, and not arranged by proximity to the mean face. In addition, subjects always viewed a subset of the 24 faces. For each occluder condition (occluded fovea vs. non-occluded fovea), there were 72 trials for each of 5 subset conditions: 1, 2, 4, 8 or 12 faces visible, which were randomly interleaved for subsets, and blocked by occlusion condition (Figure 4.3b). The faces presented on each trial were randomly drawn from the full set of 24, and as before, subjects were instructed to judge the mean emotion of the presented faces. The stimulus timing and response procedure remained identical to Experiment 1. Subjects performed 3 runs each in the non-occluded and occluded conditions for a total of 360 trials per condition. The stimuli are illustrated in Figure 4.3.

#### Analysis

Using a subset method, subjects in Experiment 2 were asked to report the perceived mean emotion of a subset of the total set of faces. Subjects viewed 1, 2, 4, 8 or 12 faces and reported the average emotion. On a trial-by-trial basis, their responses were compared to the true emotional mean of a set of 24 faces, although they were only shown a subset of faces on any given trial. For each subject, we calculated the mean of the absolute response errors, using the same method as in Experiment 1, within each of the five subset conditions (1, 2, 4, 8 and 12) and fit a line to this data using a least-squares procedure. A negative slope would indicate that subjects' performance improves as more faces are visible, and that subjects integrate multiple faces from the display. We followed a similar bootstrapping procedure to that used in Experiment 1 to determine whether the slope of the linear fit was significantly below zero. For each subject, the mean of the 1000 bootstrapped iterations for each subject, and the bootstrapped slope estimates were averaged across subjects.



**Figure 4.4:** (A) Mean of absolute response error for five subsets with foveal occlusion (blue) and without occlusion (red). We find no difference between the slopes for the two conditions; p = 0.914. There was a significant negative slope for both the non-occluded and occluded conditions (p=0.004 and p = 0.002, respectively, bootstrapped, 2-tailed) across the five subsets tested, indicating that subjects integrated more faces into their ensemble percept when available.

#### Experiment 2 Results

In Experiment 2, we measured subjects' errors in estimating the mean of the array of faces when they were only able to use a subset (1, 2, 4, 8, and 12) of the entire group to make their judgment. We found an overall decrease in mean response error with increasing set size in both the foveal occlusion and the non-occlusion conditions in this experiment (Figure 4.4). The linear fits of mean response errors as a function of subset number had significant negative slopes in both the foveal occlusion condition (slope = -0.22, p = 0.004) and in the non-occluded condition (slope = -0.24, p = 0.002; Figure 4.4). We find no significant difference between the two slopes (p = 0.91). Collapsed across both occluder conditions, compared to a subset of 4 faces, response error was significantly smaller with subsets of 8 (p = .03) and 12 (p = .016) faces. This indicates that subjects integrated information from multiple faces when available.

We also confirmed that in the one-face condition, subjects were more accurate in judging the emotion of the single face itself when they were able to foveate it in the non-occluded condition than when they could only view it in the periphery in the occluded condition (absolute mean error, 8.52 morph units in the non-occluded condition, 9.89 morph units in the occluded condition; p = 0.09). Note that this value does not reflect subjects' performance relative to the mean of the full, unseen, set of 24 (as shown in Figure 4.4) but rather it shows subjects' performance on assessing the emotion of a single face with and without foveal input. This is not surprising, but serves as a control, as it simply confirms that when available, detailed foveal information about a single face is useful.

#### **Experiment 2 Discussion**

The purpose of Experiment 2 was to determine whether subjects are able to extract ensemble information from multiple faces in the presence and absence of a foveal occluder. In particular, the subset manipulation allows us to test whether subjects' performance improves as more information is available. If subjects use multiple faces in their ensemble estimates, accuracy should improve as more faces are added. In Experiment 2, the progressive decrease in mean error, as shown by the negative slopes in each condition, indicates that subjects were able to use more faces when those faces were presented, and were able to achieve a more accurate representation of the ensemble emotion of the set, confirming previous studies (Sweeny et al., 2013; Yamanashi Leib et al., 2014). The addition of our foveal occlusion manipulation allows us to extend our demonstration of ensemble perception with emotional faces; we found no differences between the foveal occlusion and non-occlusion conditions in Experiment 1 or Experiment 2. Therefore, the efficiency of ensemble face perception does not hinge on foveal input. Different strategies may be at play with and without foveal information, and different weights might be assigned to foveally viewed faces, but the effective amount of information that subjects integrate into their ensemble percept (number of integrated faces) is consistent with and without foveal input. Our results are consistent with prior demonstrations that peripheral input is sufficient to identify single faces (Mäkelä, Näsänen, Rovamo, & Melmoth, 2001; McKone et al., 2007; Melmoth, Kukkonen, Mäkelä, & Rovamo, 2000), and that this input is also useful for ensemble processing of

groups of faces (Farzin et al., 2009; Fischer & Whitney, 2011; Haberman et al., 2009; Louie et al., 2007).

#### **General Discussion**

Ensemble perception—the ability to accurately extract the mean of a given stimulus feature from a set of objects—enables the visual system to summarize complex information in a scene in an efficient manner. Generating an ensemble representation could result from a covert process, not requiring fixation of individual objects, or an overt process of averaging foveal information across fixations. Our first experiment determined that foveal information was not necessary to perceive ensemble expression, and the second determined that, when available, multiple faces were integrated into the ensemble representation.

These results suggest that ensemble perception is not necessarily reliant on subjects foveating individual stimuli, as the inability to do so did not adversely impact performance. In contrast, (Ji et al., 2014) demonstrate that subjects weight information near the fovea more heavily when judging the average of the group. It is possible that foveal information may be given greater weight under certain circumstances—when detailed foveal information is available, when stimuli are presented briefly, or when the variability of the stimulus features is large. Even if foveal information is given greater weight than extrafoveal information under some conditions, our results demonstrate that ensemble processes are just as accurate, and just as efficient, in the absence of foveal information.

The ensemble face perception we find here, based on lower-resolution peripheral information, is consistent with previous results on ensemble perception of faces, including the fact that prosopagnosic subjects can still perceive average crowd identity (Yamanashi Leib et al., 2012), and the fact that peripherally crowded faces can contribute to an ensemble percept (Fischer & Whitney, 2011; Whitney & Levi, 2011). Evidently, ensemble perception allows humans to compensate for poor resolution—whether introduced because of retinally eccentric stimulation, crowding, perceptual deficits, or other sources. Our results hint that the visual perception of gist, in the form of ensemble information from the periphery, might be more precise and accurate than the coarse resolution limits set by acuity and crowding would suggest (Levi, 2008; Whitney & Levi, 2011). Given the effectiveness of peripheral information in generating ensemble percepts in this and previous studies, an intriguing possibility is that macular degeneration patients may show few, if any, deficits in ensemble perception, even with the total absence of foveal and parafoveal input. This will be an interesting avenue for future research.

#### Chapter 5 : Conclusion

To return to the example we discussed in Chapter 1, looking out from the stage of Wheeler Auditorium at a sea of hundreds of students, we have shown, in Chapter 2, that while the identities of the individuals who make up the crowd are crowded from our awareness, we can make accurate saccades to individuals within the crowd, and the process of doing so unlocks identifying information for the target of our saccade. While these saccades facilitated identification, this first set of experiments did not determine how much information was available presaccadically. Still within a foveal perspective, we asked how much information was available through saccadic remapping, rather than if it could simply reduce crowding. In fact, as we discussed in Chapter 3, the process of saccadic remapping acquires a full representation of the target object - in our case, a complete face – and may facilitate comparisons across different views of the world. However, this is a powerful but limited capability of the visual system; our research on the spatial tuning of saccadic remapping shows that the remapping is limited to the target object. As we have discussed from the start, peripheral input is quite detailed, and so we then asked what information could be gathered from the periphery when we stepped away from the question of individual members of the group – could an ensemble representation of the crowd be generated without foveal input? As we discussed in Chapter 4, we can generate an ensemble percept of the crowd's emotion without foveating each individual member; indeed, without foveal information at all. Peripheral input may not give us accessible representations of individual faces, but even without individuation, we can still generate an average percept of the crowd as a whole.

In closing, the research described in this dissertation shows that the visual periphery should not be discounted as a source of detailed information about the world. We may not always have access to all the information available from the periphery, but whether or not it is consciously accessible, it is a critical source of information for our perception of the world. Detailed peripheral representations mean that we can generate an average representation of the world without examining each and every object within a group, while our ability to remap object representations may allow our visual systems to more easily combine disparate views of the world into a stable unitary percept. Looking out from the stage, we can know what the overall mood of the class is – yet we can look at individual students, and still have a stable percept of the scene.

References

- Afraz, S.-R., & Cavanagh, P. (2008). Retinotopy of the face aftereffect. *Vision Research*, 48(1), 42–54. doi:10.1016/j.visres.2007.10.028
- Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, 15(3), 122–131. doi:10.1016/j.tics.2011.01.003
- Ariely, D. (2001). Seeing sets: representation by statistical properties. *Psychological Science*, 12(2), 157–162.
- Asch, S. E., & Witkin, H. A. (1948). Studies in space orientation: I. Perception of the upright with displaced visual fields. *Journal of Experimental Psychology*, 38(3), 325.
- Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision*, 9(12), 1–18. doi:10.1167/9.12.13
- Baldauf, D., Wolf, M., & Deubel, H. (2006). Deployment of visual attention before sequences of goal-directed hand movements. *Vision Research*, 46(26), 4355–4374. doi:10.1016/j.visres.2006.08.021
- Born, S., Ansorge, U., & Kerzel, D. (2012). Feature-based effects in the coupling between attention and saccades. *Journal of Vision*, *12*(11). doi:10.1167/12.11.27
- Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, 226(5241), 177–178.
- Brainard, D. H. (1997). The psychophysics toolbox. Spatial Vision, 10(4), 433–436.
- Burr, D. C., & Morrone, M. C. (2011). Spatiotopic coding and remapping in humans. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1564), 504–515. doi:10.1098/rstb.2010.0244
- Carpenter, R. H. S. (1988). *Movements of the eyes (2nd rev. & enlarged ed.)*. London, England: Pion Limited.
- Cavanagh, P., Hunt, A. R., Afraz, A., & Rolfs, M. (2010). Visual stability based on remapping of attention pointers. *Trends in Cognitive Sciences*, 14(4), 147–153. doi:10.1016/j.tics.2010.01.007
- Chakravarthi, R., & Cavanagh, P. (2009). Bilateral field advantage in visual crowding. *Vision Research*, 49(13), 1638–1646. doi:10.1016/j.visres.2009.03.026
- Chakravarthi, R., & Pelli, D. G. (2011). The same binding in contour integration and crowding. *Journal of Vision*, 11(8), 1–12. doi:10.1167/11.8.10
- Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, *43*(4), 393–404.
- Chung, S. T., Li, R. W., & Levi, D. M. (2007). Crowding between first-and second-order letter stimuli in normal foveal and peripheral vision. *Journal of Vision*, 7(2), 1:13. doi:10.1167/7.2.12
- Collins, T., Heed, T., & Röder, B. (2010). Eye-movement-driven changes in the perception of auditory space. *Attention Perception and Psychophysics*, 72(3), 736–746. doi:10.3758/APP.72.3.736
- Collins, T., Rolfs, M., Deubel, H., & Cavanagh, P. (2009). Post-saccadic location judgments reveal remapping of saccade targets to non-foveal locations. *Journal of Vision*, 9(5), 29–29. doi:10.1167/9.5.29
- Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The Eyelink Toolbox: eye

tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, & Computers : a Journal of the Psychonomic Society, Inc, 34*(4), 613–617.

- Curcio, C. A., Sloan, K. R., Kalina, R. E., & Hendrickson, A. E. (1990). Human photoreceptor topography. *The Journal of Comparative Neurology*, 292(4), 497–523. doi:10.1002/cne.902920402
- Dakin, S. C. (2001). Information limit on the spatial integration of local orientation signals. *Journal of the Optical Society of America*. a, Optics, Image Science, and Vision, 18(5), 1016–1026.
- Dakin, S. C., & Watt, R. J. (1997). The computation of orientation statistics from visual texture. *Vision Research*, *37*(22), 3181–3192.
- de Fockert, J., & Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces. *Quarterly Journal of Experimental Psychology* (2006), 62(9), 1716–1722. doi:10.1080/17470210902811249
- Deubel, H. (1995). Separate adaptive mechanisms for the control of reactive and volitional saccadic eye movements. *Vision Research*, *35*(23-24), 3529–3540.
- Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: evidence for a common attentional mechanism. *Vision Research*, 36(12), 1827–1837.
- Duhamel, J.-R., Colby, C. L., & Goldberg, M. E. (1992). The Updating of the Representation of Visual Space in Parietal Cortex by Intended Eye Movements. *Science*, 116(8), 1511–1518.
- Efron, B., & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Boca Raton, FL: CRC Press.
- Ekman, P., & Friesen, W. V. (1976). Measuring facial movement. *Environmental Psychology and Nonverbal Behavior*, *1*(1), 56–75.
- Farzin, F., Rivera, S. M., & Whitney, D. (2009). Holistic crowding of Mooney faces. Journal of Vision, 9(6), 18.1–15. doi:10.1167/9.6.18
- Fischer, J., & Whitney, D. (2011). Object-level visual information gets through the bottleneck of crowding. *Journal of Neurophysiology*, 106(3), 1389–1398. doi:10.1152/jn.00904.2010
- Grainger, J., Tydgat, I., & Isselé, J. (2010). Crowding affects letters and symbols differently. *Journal of Experimental Psychology: Human Perception and Performance*, 36(3), 673–688. doi:10.1037/a0016888
- Greene, M. R., & Oliva, A. (2009). The Briefest of Glances: The Time Course of Natural Scene Understanding. *Psychological Science*, 20(4), 464–472. doi:10.1111/j.1467-9280.2009.02316.x
- Greenwood, J. A., Bex, P. J., & Dakin, S. C. (2010). Crowding changes appearance. *Current Biology : CB*, 20(6), 496–501. doi:10.1016/j.cub.2010.01.023
- Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology*.
- Haberman, J., & Whitney, D. (2009). Seeing the mean: ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, 35(3), 718–734. doi:10.1037/a0013899
- Haberman, J., & Whitney, D. (2011). Ensemble perception: Summarizing the scene and broadening the limits of visual processing. *A Festschrift in Honor of Anne Treisman*.
- Haberman, J., Harp, T., & Whitney, D. (2009). Averaging facial expression over time.

Journal of Vision, 9(11), 1.1–13. doi:10.1167/9.11.1

- Hall, N. J., & Colby, C. L. (2011). Remapping for visual stability. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1564), 528–539. doi:10.1016/0006-8993(76)90160-8
- Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1), 29–36. doi:10.1148/radiology.143.1.7063747
- Hanley, J. A., & McNeil, B. J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, 148(3), 839– 843.
- Harrison, W. J., Mattingley, J. B., & Remington, R. W. (2013a). Eye movement targets are released from visual crowding. *Journal of Neuroscience*, 33(7), 2927–2933. doi:10.1523/JNEUROSCI.4172-12.2013
- Harrison, W. J., Retell, J. D., Remington, R. W., & Mattingley, J. B. (2013b). Visual crowding at a distance during predictive remapping. *Current Biology : CB*, 23(9), 793–798. doi:10.1016/j.cub.2013.03.050
- Helmholtz, H. (1962). *Treatise on Physiological Optics* (Vol. 3). Treatise on Physiological Optics.
- Huey, E. (1908). *The psychology and pedagogy of reading* (Vol. 9). The Macmilian Company.
- Jeon, S. T., Hamid, J., Maurer, D., & Lewis, T. L. (2010). Developmental changes during childhood in single-letter acuity and its crowding by surrounding contours. *Journal of Experimental Child Psychology*, 107(4), 423–437.
- Ji, L., Chen, W., & Fu, X. (2013). Was "seeing the mean emotion" indeed a high level analysis? *Journal of Vision*, 13(9), 591–591. doi:10.1167/13.9.591
- Ji, L., Chen, W., & Fu, X. (2014). Different Roles of Foveal and Extrafoveal Vision in Ensemble Representation for Facial Expressions (pp. 164–173). Springer.
- Jung, W. M., Bulthoff, I., Thornton, I., Lee, S. W., & Armann, R. (2013). The Role of Race in Summary Representations of Faces. *Journal of Vision*, 13(9), 861–861. doi:10.1167/13.9.861
- Kowler, E. (2011). Eye movements: the past 25 years. *Vision Research*, *51*(13), 1457–1483. doi:10.1016/j.visres.2010.12.014
- Kowler, E., & Blaser, E. (1995). The accuracy and precision of saccades to small and large targets. *Vision Research*, *35*(12), 1741–1754.
- Kowler, E., Anderson, E., Dosher, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research*, *35*(13), 1897–1916.
- Leopold, D. A., Rhodes, G., Müller, K.-M., & Jeffery, L. (2005). The dynamics of visual adaptation to faces. *Proceedings*. *Biological Sciences / the Royal Society*, 272(1566), 897–904. doi:10.1098/rspb.2004.3022
- Levi, D. M. (2008). Crowding--an essential bottleneck for object recognition: a minireview. *Vision Research*, 48(5), 635–654.
- Liberman, A., Fischer, J., & Whitney, D. (2014). Serial dependence in the perception of faces. *Current Biology*, 1–19.
- Livne, T., & Sagi, D. (2007). Configuration influence on crowding. *Journal of Vision*, 7(2), 4.1–12. doi:10.1167/7.2.4
- Livne, T., & Sagi, D. (2010). How do flankers' relations affect crowding? Journal of

*Vision*, *10*(3), 1.1–14. doi:10.1167/10.3.1

- Louie, E. G., Bressler, D. W., & Whitney, D. (2007). Holistic crowding: selective interference between configural representations of faces in crowded scenes. *Journal of Vision*, 7(2), 24.1–11. doi:10.1167/7.2.24
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281. doi:10.1038/36846
- Malania, M., Herzog, M. H., & Westheimer, G. (2007). Grouping of contextual elements that affect vernier thresholds. *Journal of Vision*, 7(2), 1.1–7. doi:10.1167/7.2.1
- Manassi, M., Sayim, B., & Herzog, M. H. (2012). Grouping, pooling, and when bigger is better in visual crowding. *Journal of Vision*, *12*(10), 1–14. doi:10.1167/12.10.13
- Maurer, D., Grand, R. L., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, 6(6), 255–260.
- Mäkelä, P., Näsänen, R., Rovamo, J., & Melmoth, D. (2001). Identification of facial images in peripheral vision. *Vision Research*, *41*(5), 599–610.
- McKelvie, S. J. (1995). Emotional expression in upside-down faces: evidence for configurational and componential processing. *The British Journal of Social Psychology / the British Psychological Society*, 34 (Pt 3), 325–334.
- McKone, E., Jeffery, L., Boeing, A., & Clifford, C. (2014). Face identity aftereffects increase monotonically with adaptor extremity over, but not beyond, the range of natural faces. *Vision Research*.
- McKone, E., Kanwisher, N., & Duchaine, B. C. (2007). Can generic expertise explain special processing for faces? *Trends in Cognitive Sciences*, 11(1), 8–15. doi:10.1016/j.tics.2006.11.002
- Melcher, D. (2005). Spatiotopic transfer of visual-form adaptation across saccadic eye movements. *Current Biology*, 15(19), 1745–1748. doi:10.1016/j.cub.2005.08.044
- Melcher, D. (2007). Predictive remapping of visual features precedes saccadic eye movements. *Nature Neuroscience*, *10*(7), 903–907. doi:10.1038/nn1917
- Melcher, D. (2009). Selective attention and the active remapping of object features in trans-saccadic perception. *Vision Research*, 49(10), 1249–1255. doi:10.1016/j.visres.2008.03.014
- Melcher, D., & Colby, C. L. (2008). Trans-saccadic perception. Trends in Cognitive Sciences, 12(12), 466–473. doi:10.1016/j.tics.2008.09.003
- Melcher, D., & Morrone, M. C. (2003). Spatiotopic temporal integration of visual motion across saccadic eye movements. *Nature Neuroscience*, 6(8), 877–881. doi:10.1038/nn1098
- Melmoth, D. R., Kukkonen, H. T., Mäkelä, P. K., & Rovamo, J. M. (2000). The effect of contrast and size scaling on face perception in foveal and extrafoveal vision. *Investigative Ophthalmology & Visual Science*, 41(9), 2811–2819.
- Murdock, B. B. (1965). Signal-detection theory and short-term memory. *Journal of Experimental Psychology*, 70(5), 443–447.
- Myczek, K., & Simons, D. J. (2008). Better than average: alternatives to statistical summary representations for rapid judgments of average size. *Perception and Psychophysics*, 70(5), 772–788. doi:10.3758/PP.70.5.772
- Nakamura, K., & Colby, C. L. (2002). Updating of the visual representation in monkey striate and extrastriate cortex during saccades. *Proceedings of the National Academy of Sciences of the United States of America*, 99(6), 4026–4031.

doi:10.1073/pnas.052379899

- Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences*, 11(12), 520–527. doi:10.1016/j.tics.2007.09.009
- Pajak, M., & Nuthmann, A. (2013). Object-based saccadic selection during scene perception: Evidence from viewing position effects. *Journal of Vision*, 13(5). doi:10.1167/13.5.2
- Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, 4(7), 739–744. doi:10.1038/89532
- Parks, N. A., & Corballis, P. M. (2008). Electrophysiological correlates of presaccadic remapping in humans. *Psychophysiology*, 45(5), 776–783. doi:10.1111/j.1469-8986.2008.00669.x
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, 10(4), 437–442.
- Pelli, D. G. (2008). Crowding: a cortical constraint on object recognition. *Current Opinion in Neurobiology*, 18(4), 445–451. doi:10.1016/j.conb.2008.09.008
- Pelli, D. G., & Tillman, K. A. (2008). The uncrowded window of object recognition. *Nature Neuroscience*, *11*(10), 1129–1135.
- Pelli, D. G., Palomares, M., & Majaj, N. J. (2004). Crowding is unlike ordinary masking: distinguishing feature integration from detection. *Journal of Vision*, 4(12), 1136– 1169.
- Piazza, E. A., Sweeny, T. D., Wessel, D., Silver, M. A., & Whitney, D. (2013). Humans use summary statistics to perceive auditory sequences. *Psychological Science*, 24(8), 1389–1397. doi:10.1177/0956797612473759
- Poliak, S. (1941). The retina. The University of Chicago Press, Chli-Aco, Illinois, 1, 41.
- Remington, R. W. (1980). Attention and saccadic eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, 6(4), 726–744.
- Rensink, R. A. (2002). Change detection. Annual Review of Psychology, 53(1), 245-277.
- Rhodes, G., Evangelista, E., & Jeffery, L. (2009). Orientation-sensitivity of face identity aftereffects. *Vision Research*, 49(19), 2379–2385. doi:10.1016/j.visres.2009.07.010
- Rhodes, G., Jeffery, L., Clifford, C. W. G., & Leopold, D. A. (2007). The timecourse of higher-level face aftereffects. *Vision Research*, 47(17), 2291–2296. doi:10.1016/j.visres.2007.05.012
- Rhodes, G., Jeffery, L., Watson, T. L., Jaquet, E., Winkler, C., & Clifford, C. W. G. (2004). Orientation-contingent face aftereffects and implications for face-coding mechanisms. *Current Biology*, 14(23), 2119–2123. doi:10.1016/j.cub.2004.11.053
- Rolfs, M., Jonikaitis, D., Deubel, H., & Cavanagh, P. (2010). Predictive remapping of attention across eye movements. *Nature Neuroscience*, 14(2), 252–256. doi:10.1038/nn.2711
- Rutherford, M. D., Chattha, H. M., & Krysko, K. M. (2008). The use of aftereffects in the study of relationships among emotion categories. *Journal of Experimental Psychology: Human Perception and Performance*, 34(1), 27–40. doi:10.1037/0096-1523.34.1.27
- Sayim, B., Westheimer, G., & Herzog, M. H. (2008). Contrast polarity, chromaticity, and stereoscopic depth modulate contextual interactions in vernier acuity. *Journal of Vision*, 8(8), 1–9.

- Sayim, B., Westheimer, G., & Herzog, M. H. (2010). Gestalt factors modulate basic spatial vision. *Psychological Science*, 21(5), 641–644. doi:10.1177/0956797610368811
- Schneider, W. X., & Deubel, H. (1995). Visual attention and saccadic eye movements: Evidence for obligatory and selective spatial coupling. *Studies in Visual Information Processing*, 6, 317–324.
- Schotter, E. R., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. Attention Perception and Psychophysics, 74(1), 5–35. doi:10.3758/s13414-011-0219-2
- Sherman, A. M., Evans, K. K., & Wolfe, J. M. (2012). The gist of the organized is more precise than the gist of the random. *Journal of Vision*, 12(9), 841–841. doi:10.1167/12.9.841
- Sommer, M. A., & Wurtz, R. H. (2008). Brain circuits for the internal monitoring of movements. *Annual Review of Neuroscience*, 31, 317–338. doi:10.1146/annurev.neuro.31.060407.125627
- Strasburger, H., Rentschler, I., & Juttner, M. (2011). Peripheral vision and pattern recognition: A review. *Journal of Vision*, 11(5), 13–13. doi:10.1167/11.5.13
- Subramanian, J., & Colby, C. L. (2014). Shape selectivity and remapping in dorsal stream visual area LIP. *Journal of Neurophysiology*, 111(3), 613–627. doi:10.1152/jn.00841.2011
- Suzuki, S., & Cavanagh, P. (1998). A shape-contrast effect for briefly presented stimuli. Journal of Experimental Psychology: Human Perception and Performance, 24(5), 1315–1341.
- Sweeny, T. D., & Whitney, D. (2014). Perceiving Crowd Attention: Ensemble Perception of a Crowd's Gaze. *Psychological Science*, 25(10), 1903–1913. doi:10.1177/0956797614544510
- Sweeny, T. D., Haroz, S., & Whitney, D. (2013). Perceiving group behavior: Sensitive ensemble coding mechanisms for biological motion of human crowds. *Journal of Experimental Psychology: Human Perception and Performance*, 39(2), 329–337. doi:10.1037/a0028712
- Valentine, T. (1988). Upside-down faces: a review of the effect of inversion upon face recognition. *British Journal of Psychology (London, England : 1953)*, 79 (Pt 4), 471–491.
- van Koningsbruggen, M. G., & Buonocore, A. (2013). Mechanisms behind perisaccadic increase of perception. *Journal of Neuroscience*, 33(28), 11327–11328. doi:10.1523/JNEUROSCI.1567-13.2013
- Vlaskamp, B., & Hooge, I. (2006). Crowding degrades saccadic search performance. *Vision Research*, *46*(3), 417–425.
- Wallace, J. M., & Tjan, B. S. (2011). Object crowding. *Journal of Vision*, 11(6), 1–17. doi:10.1167/11.6.19
- Watamaniuk, S. N., & Sekuler, R. (1992). Temporal and spatial integration in dynamic random-dot stimuli. *Vision Research*, *32*(12), 2341–2347.
- Webster, M. A., Kaping, D., Mizokami, Y., & Duhamel, P. (2004). Adaptation to natural facial categories. *Nature*, 428(6982), 557–561.
- Whitney, D., & Levi, D. M. (2011). Visual crowding: a fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences*, *15*(4), 160–168.

doi:10.1016/j.tics.2011.02.005

- Wolfe, B. A., & Whitney, D. (2014). Facilitating recognition of crowded faces with presaccadic attention. *Frontiers in Human Neuroscience*, 8, 103. doi:10.3389/fnhum.2014.00103
- Wurtz, R. H. (2008). Neuronal mechanisms of visual stability. Vision Research, 48(20), 2070–2089. doi:10.1016/j.visres.2008.03.021
- Wurtz, R. H., Joiner, W. M., & Berman, R. A. (2011). Neuronal mechanisms for visual stability: progress and problems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1564), 492–503. doi:10.1098/rstb.2010.0186
- Yamanashi Leib, A., Fischer, J., Liu, Y., Qiu, S., Robertson, L., & Whitney, D. (2014). Ensemble crowd perception: A viewpoint-invariant mechanism to represent average crowd identity. *Journal of Vision*, 14(8). doi:10.1167/14.8.26
- Yamanashi Leib, A., Puri, A. M., Fischer, J., Bentin, S., Whitney, D., & Robertson, L. (2012). Crowd perception in prosopagnosia. *Neuropsychologia*, 50(7), 1698–1707. doi:10.1016/j.neuropsychologia.2012.03.026
- Yang, J.-W., Yoon, K. L., Chong, S. C., & Oh, K. J. (2013). Accurate but pathological: Social anxiety and ensemble coding of emotion. *Cognitive Therapy and Research*, 37(3), 572–578.
- Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, 81(1), 141.