# Can perceptual averaging *really* occur in the absence of change localization?

A Thesis

Submitted In Partial Fulfillment of the Requirements

for the Degree of

Bachelor of Science (Honours) in Psychology

University of Regina

by

Candice Giesinger

Regina, Saskatchewan

April 24, 2017

Abstract

Noticing the location of an object that causes a change to the mean of a set relies on the ability to determine the mean of the set, and detect that a change has occurred (Rensink, 2002). Previous research suggests that people are able to retain information about the mean emotion of a set of faces even when they are unsure which items changed between the two sets (Haberman & Whitney, 2011). Subjects in that study, however, could use a strategy of localizing the most emotionally extreme face in the set to reliably determine the correct response in the mean discrimination task. In the present study, the utility of this strategy was eliminated. Subjects completed 4 blocks of trials consisting of 48 trials per block. On each trial, subjects viewed two consecutive displays of faces contained within circles. Four items increased (or decreased) in size or emotional intensity. In Experiment 1, subjects first determined whether average size or emotion increased or decreased from the first display to the second, then localized one of the four changed items. In Experiment 2, the order of responding was reversed. The results suggest that when performing both a mean discrimination and localization task, subjects use their knowledge of which stimulus in the set changed to guide their response on the mean discrimination task. Focusing attention to a local region of a display prevents the global distribution of attention necessary for perceptual averaging (Chong & Treisman, 2003). Thus, averaging is not possible when change detection fails.

*Keywords*: change localization, mean discrimination, statistical summary, size, emotion

# Can perceptual averaging *really* occur in the absence of change localization?

Sets of objects are commonplace in our daily lives. Information retained from sets can vary, from highly specific information about a particular item of interest to a more general representation of the set as a whole (Ariely, 2001). The attention devoted to the particular set can be focused on a single object or spread to a wider area. When attention is focused on a single object, local properties of that object can be determined, whereas diffusing attention over a wider area allows the observer to represent global properties of the set (Chong & Treisman, 2005). Statistical summary representations (SSRs) are mental representations of the statistical properties of a set, such as the mean, and are said guide behaviour more efficiently than representing each item individually (Attarha, Moore, & Vecera, 2014). SSRs would allow one, for example, to determine the mean grape size at the supermarket to select the bunch with the largest fruit, or to determine the average emotion of people in a classroom to decide whether or not to stop teaching for the day.

SSRs could also contribute to detection of changes in sets. Often, it is necessary to notice a change between two different sets, and doing so can provide relevant real-world information, such as allowing a driver to notice that the car in front of her has changed lanes. Although most research on SSRs uses simple stimuli such as circles or lines, to date, research on the role of SSRs in change detection has been done exclusively with faces. This presupposes that sets of faces and circles are processed in a similar manner; however, no study has tested this assumption directly by comparing the processing of both faces and circles at the same time, despite clear evidence that individual circles and faces are processed quite differently.

Recently, a study was published that found that subjects were able to retain information pertaining to a change in the average emotion of a set of faces, without being able to retain information about which items in the set had changed (Haberman & Whitney, 2011). The current study combines mean discrimination and change localization tasks using both faces and circles in an attempt to better understand the mechanisms that subserve this ability. The text that follows examines relevant SSR literature, as well as pertinent change detection literature. It will then show how the two paradigms were combined into a single study before moving into the specific methods and implications of the current study.

**Properties of Statistical Summary Representations**

Many properties of sets of stimuli are accurately captured by SSRs, suggesting that the computing of SSRs is done through a general mechanism. SSRs can accurately represent lower order properties of a set including orientation, speed, and length, as well as higher order properties such as average emotion and gender (Attarha et al., 2014). Why might sets of items be represented differently from the individual items that comprise them? One function of SSRs is to decrease the processing load in unattended areas and in the periphery (Attarha et al., 2014), allowing the entire set to be summarized within a single representation rather than as a distributed collection of individual representations.

Several findings support the idea that precise information is lost in favour of a more general representation. One piece of evidence comes from contrasting performance in an averaging task, where subjects are asked to report the average of some property of a set, with performance on a member identification task, where subjects are asked to judge whether a particular item previously appeared in a set. Subjects who are asked to recall specific items differing along a single dimension (e.g., size) from a set in a member identification task are unable to perform better than expected by chance (Ariely, 2001; Attarha, et al., 2014); however, subjects are typically very accurate in reporting the mean of the set. In addition, previous research conducted with sets has suggested that size information about the individual exemplars is not accessible (Neumann, Schweinberger, & Burton, 2013). Neumann and colleagues argue that in highly variable sets, the amount of variability in the set may limit the maximum number of individual exemplars that can be encoded, or retained, forcing the visual system to summarize the sets according to their statistical properties. Alternatively, the visual system might have the flexibility to represent the set either as a summary representation or as a collection of individual representations, if such individual representations are required for the task.

The representation of sets is not limited to basic features such as size or colour. When presented with sets of complex stimuli such as faces, subjects are able to extract the mean emotionality of the faces in the set, and are more likely to indicate that they saw a stimulus that corresponds to the average rather than a stimulus that was actually presented (Neumann et al., 2013). This would suggest that subjects might be encoding sets of faces using the same mechanism used to compute SSRs for lower level features like size or colour (de Fockert & Wolfenstein, 2009).

In sum, SSRs are used to represent sets of many stimulus types, including both simple stimuli like circles and more complex stimuli like faces. Precise information from a set is typically lost, but general characteristics such as mean size remain, and are highly accurate. Individual exemplars in a set may be encoded, but are not accessible. When sets of faces are presented, subjects are able to determine mean emotionality, and when the faces are unique individuals, subjects form an average of the presented identities. Because sets of items are likely encoded using SSRs, a question that naturally follows is whether computing SSRs can facilitate the detection of changes between two different sets. Before addressing this question, a brief overview of the relevant change detection literature is provided.

**Factors Influencing Change Detection**

Change detection is defined as the ability to detect changes in an object or scene (Simons & Levin, 1997). In a change detection experiment, subjects are shown a scene (A) followed by a blank interval, which is followed by a modified version (A′) of the original scene. Encoding individual details of a scene would require attending to each aspect of the scene, overwhelming the limited attentional capacity of the observer (Simons & Levin, 1997). To overcome these limitations, people tend to instead encode the gist of a scene, which usually remains the same following a change. If subjects have encoded only the gist without encoding individual details, change detection should fail. Nevertheless, encoding the gist is usually sufficient, as the smaller details are irrelevant for noticing changes in most day-to-day situations (Haberman & Whitney, 2001; Simons & Levin, 1997). Thus, although capacity limitations prevent detailed encoding of a visual scene, encoding of its gist is usually sufficient for most tasks.

It is important to note the distinction between change detection and change localization. Change detection, as mentioned above, is the ability to notice a change has occurred. Change localization, on the other hand, is the ability to identify which aspect(s) of a scene have changed (Ball & Busch, 2015). It has been argued that change detection can occur without knowledge of the location of the change when this change occurs outside of the focus of attention; however, in all cases, a change must be detected before it can be localized (Ball & Busch, 2015). Yet, the encoding of gist suggests that very little visual information is retained from one view to the next, and once the eyes move to encode a new view, access to visual details of the previous view are lost (Simons & Levin, 1997). If detailed representations are not encoded, then how are changes successfully localized when change detection succeeds? Research suggests that the detection of change depends on interpreting motion signals across the retina (Simons & Levin, 1997;

Rensink, 2002). When subjects are asked to compare displays in which a blank display intervenes, even if there is only one change, the appearance of A′ floods the brain with motion signals making it difficult to detect the change. Simply put, a change is not detected because the eye is processing change signals from every location (Simons & Levin, 1997).

In addition to the type of change made to the display, where the subject focuses their attention also plays a role in change detection (Ball & Busch, 2015; Simons & Levin, 1997). Changes to the display that are located in the center of the display are noticed more readily than those on the periphery of the scene, suggesting that attention is focused more often or more quickly to central objects in the scene, which allows for faster change detection. Despite this, changes to central objects are not always noticed. Whether or not a change is expected also influences the ability to detect a change. Even if a subject is expecting a change, and intentionally devotes attention to noticing it, change blindness still frequently occurs (Rensink, 2002). Taken together, this evidence suggests that attention is necessary, but insufficient, for change detection.

In sum, the visual system stores very little information from one view to the next. Changes made to the centre of the display are noticed more often than changes on the periphery. Attention is essential for change detection to occur, and change blindness is not due to a failure to focus attention on the objects prior to the change. Subjects typically encode the gist of a scene, which can result in change blindness if the subject only encoded the gist, as the gist tends to remain the same between views.

**Combining Change Detection and SSRs**

Given that both SSRs and change detection appear to rely on more global representations of scenes at the expense of representing individual items, a question that naturally follows is whether SSRs can facilitate change detection. To date, only one study has combined both the change detection and SSR paradigms. Haberman and Whitney (2011) examined whether subjects were able localize a change between two sets of sequentially presented faces, while also judging the mean emotion of the faces. Subjects viewed two successive sets of emotionally varying faces and performed both mean discrimination and change localization tasks on the same trial. In the mean discrimination task, subjects indicated which of the two sets of images had on average the happier expression. In the change localization task, subjects indicated where in the set a change in emotionality occurred. The set of gray-scaled faces varied in increments from happy, to neutral, to sad, and back to happy again, creating a continuum of emotion between exemplars with different emotions. The sets consisted of four occurrences of four unique images making 16 faces per set. In the second display, four of the most emotionally extreme faces in the first display (either the happiest or saddest) changed to the other emotional extreme.

The results of Haberman and Whitney's (2011) study suggested that subjects were able to derive sufficient information from two or three faces on each trial to be able to do the task, as performance on the change localization task was significantly better than chance. Performance on the change detection task was also significantly above chance levels, suggesting subjects had information about the average emotion of each set. Excluding trials on which change localization was successful, examining the change detection task again revealed that performance remained significantly above chance: despite being unsuccessful in localizing the change, observers were able to notice the direction in which the change in emotionality had occurred. Change detection performance was lower when change localization failed, but remained above chance.

The results of Haberman and Whitney's (2011) study may have been caused by the methodology implemented in this study. By changing the four most emotionally extreme faces between sets, subjects would be able to use several strategies that do not depend on encoding the mean to perform well on the task. One such strategy would be attending to the most emotionally extreme face only, as the changes that occurred to that stimulus would be indicative of how the average emotion of the set as a whole would change. This strategy, however, predicts that subjects would be able to localize the change every time they succeeded on the change detection task, which was not the case. To account for this, it could instead be assumed that subjects are, mistakenly, not attending to the most emotionally extreme face. For example, they may attend to the second most emotionally extreme face, which would not change between displays. In this situation, the subject is required to make a random guess as to which stimulus changed; but, they can make an educated guess about the direction of the overall emotional change to the set based on the emotion of the face to which they previously attended. This strategy would yield a pattern of results corresponding to the change detection success with change localization failure that occurred in Haberman and Whitney's (2011) study, even if subjects did not actually compute the mean properties of the presented set. As such, it is unclear whether change detection in this task depended on a calculation of average emotion. The purpose of the current study is to determine if change detection can succeed when localization fails, controlling for strategies that would allow subjects to circumvent averaging altogether.

## Experiment 1

### Method

**Subjects.** 30 subjects, 7 of which were left-handed, consisting of 27 females, and 3 males (*M*: 21.3 years, *SD*: 2.2), from the University of Regina Psychology Participant Pool completed this experiment. The Participant Pool consists of students enrolled in 100 and 200 level psychology courses at the University of Regina. Subjects received course credit for their participation, and were required to have normal or corrected-to-normal vision.

**Stimuli and Apparatus.** Stimuli were displayed in black on a white background on an LCD monitor with a display resolution of 1920 X 1200 pixels, with a vertical refresh rate of 60 Hz. Stimuli were presented on an imaginary grid which consisted of 16 non-overlapping boxes, centered on fixation. Each stimulus was presented in the center of each imaginary box. Stimuli were selected at random with replacement from the full set of stimuli with the constraint that no more than four repetitions of any one stimulus could occur in any display. There was no restriction on the number of unique stimuli that could appear in each display, meaning that there was the potential for a display to consist of as many as 10 unique stimuli. The mean size or emotion of each display varied from trial to trial depending on the stimuli randomly selected by the computer to comprise the displays. The sizes of the circles in the size condition ranged from 1.05 to 4.64 degrees of visual angle. For the emotion condition, circle size was constant at 4.19 degrees of visual angle.

Stimuli consisted of grey-scaled faces contained within circles so that there was no blank space between the edge of the face and the edge of the circle (Figure 1). Facial expressions were created by morphing photos of faces (Tottenham et al., 2009) displaying happy or sad expressions with neutral photos. Expressions ranged from 100% happy to 100% neutral to 100% sad with 10% increments of increased neutrality between expressions (e.g., 100H/0N, 90H/10N… 0H/100N… 10N/90S, 0N/100S). This increment level was chosen as a compromise between having enough stimuli to ensure a range of variability of expression within each display, and the ability to distinguish the expressions as different from each other. Six different

Caucasian female identities were used, resulting in 180 different stimuli. Only one identity was presented per trial.

There were eight unique trial types repeated 6 times to create a block, resulting in 48 trials per block. Subjects completed four unique blocks, two judging size and two judging emotion. The emotion block instructions asked subjects to select which of the two displays was happier or sadder, depending on the condition. The size block instructions asked subjects to select which of the two displays was larger or smaller, depending on the condition. All subjects completed all conditions, and the order of conditions was counterbalanced across subjects (24 subjects were required for a full counterbalance). Trials could either be overall happy or sad, for the emotion trials, or large or small, for the size trials. For the emotion trials, each set was either overall happy, or overall sad, with varying amounts of neutrality. Each display contained a full range of emotion and size.

Each trial consisted of two unique displays. To create these displays, four stimuli within either the upper or lower half of the range of stimulus intensities were randomly selected to differ in intensity across the first and second displays. The stimuli chosen to differ across the two displays were equally likely to be chosen from the lower and upper intensities. On half of size trials, the stimuli chosen to change regressed to the middle size within the range from which they were selected. Specifically, stimuli drawn from the smaller range of sizes changed to the size corresponding to the midpoint of this range (1.89 degrees of visual angle), and stimuli drawn from the larger range of sizes changed to the size corresponding to the midpoint of this range (3.65 degrees of visual angle). For the emotion trials, all stimuli that changed were presented as 100% neutral stimuli in the following display. On half of trials, the order of displays was reversed, such that the second display was more variable than the first.

**Procedure.** Trials began with a fixation cross at the centre of the display measuring .34 X .34 degrees of visual angle that remained in view until the subject indicated they were ready to begin the trial by pressing the space bar. After the response, a blank screen was presented for 300 milliseconds (ms). Each trial consisted of two displays that were presented for 1000 ms each, with a 500 ms blank screen intervening between the two displays. Immediately following the second display, the response screen was presented, which remained until the subject responded. The response screen consisted of a letter grid which corresponded to the position of the previously presented stimuli.[1] The letters on the grid were located in the center of each of the 16 boxes contained in the imaginary grid described above. Below the letter grid, about one-sixth of the way from the bottom of the screen was the mean discrimination task prompt, which read: "Is 1 or 2 ____?" where the blank would read *happier*, *sadder*, *smaller*, or *larger*, depending on the condition. Both the letter grid and the mean discrimination prompt were present on the screen until the subject responded. Subjects were required to respond to the mean discrimination prompt first, which then disappeared from the screen. After responding to the mean discrimination prompt, subjects completed the localization task in which they were required to select the letter from the grid corresponding to the location of one of the changed items. Once a response was made, the grid disappeared and was replaced by the fixation cross, ending the trial.

---

[1] There was small typing error in the letter grid, which resulted in two letter "i"'s appearing. The second "i" was where the letter "l' should have been. As such, it could not be accurately determined which letter the subject was indicating when they responded with either of these letters. To compensate for this, all trials in which subjects responded with an "i" or an "l" were excluded from analysis.

In addition to the four experimental blocks, subjects also completed a block of eight practice trials in one of the four tasks, randomly chosen by the computer. The practice trials were identical to the experimental trials, with the exception that feedback was provided on the practice trials. After both the mean discrimination task and the localization task responses had been gathered from the subject, feedback was provided for 1500 ms. To the left of fixation, feedback on the mean discrimination task was given. To the right of fixation, feedback on the localization task was given. The colour of the text was determined by the accuracy of the response, with green text for correct responses and red text for incorrect responses.

**Design.** Stimulus type (emotion and size) and task type (mean discrimination and localization) were combined factorially, resulting in a 2 x 2 design. The order of stimulus type was randomly determined by the computer, and accuracy was measured on both tasks. Thus, the independent variables were stimulus type and task type, with accuracy as the dependent variable. In analyzing the data, particular attention was given to performance on the mean discrimination task when the localization task was unsuccessful.

## Results

**Mean Discrimination Task.** A two-way repeated measures analysis of variance (ANOVA) was conducted to compare the effects of localization performance (correct vs. incorrect localization) and stimulus type (circles vs. faces) on averaging performance. There was no statistically significant difference between the stimulus types, or interaction between stimulus type and localization performance, $F(1,29)=.21$, $p > .05$, $\eta_p^2=.007$. Accuracy on the mean discrimination task was higher following a correct localization response (*M*: 71.1%) than an incorrect localization response (*M*: 53.0%). This difference was statistically significant, collapsing over stimulus type, $F(1, 29)=87.92$, $p < .001$, $\eta_p^2=.75$. The results of a one-sample t-test, summarized in Figure 2, showed that subjects' performance on the mean discrimination task (*M*=71.65%, *SD*: 13.75%) was statistically significantly above chance levels (50%), when they were successful on the localization task, $t(29)=8.22$, $p < .001$, 95% *CI* [0.16, 0.26]. The results of a one-sample t-test also showed that subjects' performance on the mean discrimination task (*M*=52.65%, *SD*: 6.25%) was statistically significantly different from chance (50%), when they were unsuccessful on the localization task, $t(29)=2.32$, $p < .05$, 95% *CI* [0.00, 0.05].

**Localization Task.** A two-way repeated measures ANOVA was conducted to compare the effects of the mean discrimination task (correct vs. incorrect mean discrimination) and stimulus type (circles vs. faces) on localization performance. There was no statistically significant difference between the stimulus types, or interaction between stimulus type and performance on the mean discrimination task, $F(1,29)=2.03$, $p > .05$, $\eta_p^2=.07$. Accuracy on the localization task was higher following a correct mean discrimination response (*M*: 32.3%) than following an incorrect mean discrimination response (*M*: 17.8%), with an increase of 7% accuracy from chance levels for correct trials and a decrease of 7% accuracy from chance levels for incorrect trials. This difference between the means was statistically significant, collapsing over stimulus type, $F(1,29)=75.34$, $p < .001$, $\eta_p^2=.72$. The results of a one-sample t-test, summarized in Figure 2, showed that subjects' performance on the localization task (*M*=32.25%, *SD*: 8.87%) was statistically significantly above chance levels (25%), when they were successful on the mean discrimination task, $t(29)=4.48$, $p < .001$, 95% *CI* [0.04, 0.12]. The results of a one-sample t-test also showed that the subjects' performance on the localization task (*M*=17.81%, *SD*: 8.31%) was statistically significantly below chance levels (25%), when they were unsuccessful on the mean discrimination task, $t(29)=4.74$, $p < .001$, 95% *CI* [-0.10, -0.04].

**Discussion**

Performance on both the mean discrimination and localization tasks was significantly above chance levels when subjects were correct on the other task. Performance on the mean discrimination task remained above chance levels even when the localization task was unsuccessful. Subjects appear to have reliable information about the mean emotion of the set of faces, even when they are unsure which items changed between the displays. These results are consistent with those found by Haberman and Whitney (2011). The results suggest that subjects are able to determine a change in the average expression of a set of faces, even when they are unable to identify which expressions changed.

Performance on the localization task was above chance levels when the mean discrimination task was successful and below chance levels when it was unsuccessful. This suggests that performance on the localization task is dependent on performance on the averaging task. This is especially apparent given the 7% increase in accuracy above chance levels when the mean discrimination task was successful, and the 7% decrease in accuracy below chance levels when the mean discrimination task was unsuccessful. One explanation for these results is that subjects are highly accurate when, purely by chance, they happen to be attending to one of the stimuli in the set that changes between displays, and use this information to answer both the mean discrimination and localization prompts. This pattern also implies that if subjects are unsuccessful on the averaging task, it is because they mistakenly thought that a stimulus changed in a particular way (even though in actuality it did not change at all), and then selected the answer to the prompt that matched their mistaken perception.

Although the results are consistent with the suggestion that subjects can determine a change in average size and emotion in the absence of localizing changes to single items, another possibility arises from the fact that in Experiment 1, subjects were required to respond to the mean discrimination prompt before being able to respond to the localization prompt. It is possible that subjects are correctly localizing the change, but are responding incorrectly after they have completed the averaging task because they have simply forgotten which item changed. This would artificially inflate their mean discrimination accuracy for incorrect localization trials, as these trials would be counted as a success on the mean discrimination task in the absence of successful localization, even though subjects really did notice which item in the set changed. This explanation would account for the pattern of increased and decreased localization performance that accompanied correct or incorrect performance on the mean discrimination task, respectively.

To test the possibility that subjects are correctly localizing the change, but forgetting the location by the time they respond to the localization prompt, a second experiment was conducted that reversed the order of responding. By having the localization task first, the possibility of the subjects forgetting the location of a correctly localized change is reduced. Thus, it would be expected that subjects are only incorrect on the localization task when they truly do not localize the change. If they are still able to determine the direction of change in average size and emotion on such trials, this would provide strong evidence that localization of change is not necessary to determine changes in the average of a property of a set.

## Experiment 2

### Method

30 right-handed subjects, consisting of 24 females and 6 males (*M*: 20.5 years, *SD*: 0.4), from the University of Regina Psychology Participant Pool completed this experiment. The Participant Pool consists of students enrolled in 100 and 200 level psychology courses at the University of Regina. Subjects received course credit for their participation, and were required to

have normal or corrected-to-normal vision. In Experiment 2 all methods were identical to those in Experiment 1, with the exception that only the letter grid appeared after the second display, and once the subjects responded in the localization task, the prompt for the mean discrimination task appeared and remained until response.

**Results**

**Mean Discrimination Task.** A two-way repeated measures ANOVA was conducted to compare the effects of localization performance (correct vs. incorrect localization) and stimulus type (circles vs. faces) on averaging performance. There was no statistically significant difference between the stimulus types, $F(1,29)=2.63$, $p > .05$, $\eta_p^2=.08$. There was a statistically significant interaction between stimulus type and the effects of localization performance on mean discrimination performance. This indicates that the difference in performance on the localization task between successful and unsuccessful mean discrimination task performance was larger for size trials (76.50% vs. 49.74%) than for emotion trials (67.48% vs. 51.96%), $F(1,29)=7.59$, $p < .05$, $\eta_p^2=.21$. The results of a one-sample t-test, summarized in Figure 2, showed that subjects' performance for emotion trials on the mean discrimination task (*M*=51.96%, *SD*: 8.80%) was not statistically significantly different from chance (50%), when they were unsuccessful on the localization task, $t(29)=1.22$, $p > .05$, 95% *CI* [-0.01, 0.05]. The same was true for the size trials, in that performance on the mean discrimination task (*M*=49.74%, *SD*: 5.75%) was not statistically significantly different from chance, when subjects were unsuccessful on the localization task, $t(29)=.25$, $p > .05$, 95% *CI* [-0.02, 0.02].

**Localization Task.** A two-way repeated measures ANOVA was conducted to compare the effects of the mean discrimination task (correct vs. incorrect mean discrimination) and stimulus type (circles vs. faces) on localization performance. There was no statistically significant difference between the stimulus types, $F(1,29)=0.71$, $p > .05$, $\eta_p^2=.02$. There was a statistically significant interaction between stimulus type and the effects of mean discrimination performance on localization performance. This indicates that the difference in performance on the mean discrimination task between successful and unsuccessful localization task performance was larger for size trials (37.77% vs. 16.06%) than for emotion trials (32.35% vs. 19.15%), $F(1,29)=7.85$, $p < .05$, $\eta_p^2=.21$. The results of a one-sample t-test, summarized in Figure 2, showed that the subjects' performance for the emotion trials on the localization task ($M$=19.15%, $SD$: 7.99%) was statistically significantly below chance levels (25%), when they were unsuccessful on the mean discrimination task, $t(29)=4.01$, $p < .001$, 95% $CI$ [-0.09, -0.03]. The same pattern of results was also found for the size trials, as the results of a one-sample t-test showed that the subjects' performance on the localization task ($M$=16.06%, $SD$: 8.50%) was statistically significantly below chance levels (25%), when subjects were unsuccessful on the mean discrimination task, $t(29)=5.76$, $p < .001$, 95% $CI$ [-0.12, -0.06].

**Discussion**

Similar to Experiment 1, for both the mean discrimination task and the localization task, performance was significantly above chance when the subject was correct on the other task. This suggests that success on one task should correspond to successful performance on the other. Consistent with the results of Experiment 1, performance on the localization task was statistically below chance levels when the mean discrimination task was unsuccessful, and above chance levels when the mean discrimination task was successful. This indicates that localization performance is dependent on mean discrimination performance.

Notably, performance on the mean discrimination task was statistically different from chance only when the localization task was successful. These results are not consistent with those found by Haberman and Whitney (2011). When subjects are forced to respond to the localization task first, the ability to successfully complete the mean discrimination task when localization fails disappears. The results suggest that when performing both an averaging and localization task, subjects are using their knowledge of which stimulus in the set changed to guide their performance on the averaging task. There is no evidence from the present study that averaging is possible when change localization fails. Although this is inconsistent with Haberman and Whitney's (2011) finding, this result is not surprising, as averaging tasks require global attention and change localization promotes local attention (Chong & Treisman, 2003).

## General Discussion

Taken together, the results of Experiment 1 and 2 indicate that subjects are not able to extract information about the mean of a property of a set of objects when required to complete both mean discrimination and localization tasks. It is important to note that the same results held for both stimulus types, indicating that these trends are not unique to one stimulus type. The results of Experiment 2 revealed that subjects are using their knowledge of which items in the set changed to respond to the mean discrimination prompt. The fact that performance was at chance levels on the mean discrimination task when the localization task was unsuccessful suggests that subjects have no reliable information about the mean of the set. The two tasks promote different attentional strategies, with the mean discrimination task promoting global attention and the localization task promoting local attention. It is likely that this incongruence of attention is what is preventing subjects from performing well on both tasks at the same time, as it has been shown that subjects are able to reliably determine the mean of various stimuli (Attarha et al., 2014) when they are not required to also attend to a single stimulus in the set (cf. de Fockert & Marchant, 2008).

As with any research, there were limitations with the current study. One limitation is that while it is suspected that different attentional demands are causing the results, it was beyond the scope of this study to test this explanation. To further explore the results of this study, future research should examine under what conditions subjects are able to determine the mean properties of a set when combined with a localization task. One way in which this could be implemented would be to reduce the overall set size, in order to reduce the disparity in the demands of the global and local deployment of attention that is promoted in each task. Previous research (Haberman & Whitney, 2011) has shown that subjects use information from only two or three items of a presented set, and so reducing the number of presented items should not impact the difficulty of the task, but will reduce the need for global attention. If decreasing set size yields improved performance on the mean discrimination task with unsuccessful localization, this would be evidence suggesting that mean discrimination of a set of items can occur without change localization but only with a limited number of items demanding attention.

References

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science, 12*, 157-162. doi:10.1111/1467-9280.00327

Attarha, M., Moore, C. M., & Vecera, S. P. (2014). Summary statistics of size: Fixed processing capacity for multiple ensembles but unlimited processing capacity for single ensembles. *Journal of Experimental Psychology. Human Perception and Performance*, *40*, 1440-1449. doi:10.1037/a0036206

Ball, F., & Busch, N. A. (2015). Change detection on a hunch: Pre-attentive vision allows “sensing” of unique feature changes. *Attention, Perception, & Psychophysics, 77*, 2570-2588. doi:10.3758/s13414-015-0963-9

Chong, S. C., & Treisman, A. (2005). Attentional spread in the statistical processing of visual displays. *Perception & Psychophysics*, *67*, 1-13. doi:10.3758/BF03195009

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research, 43,* 393– 404. doi:10.1016/S0042-6989(02)00596-5

de Fockert, J. W., & Marchant, A. P. (2008). Attention modulates set representation by statistical properties. *Perception & Psychophysics*, *70*(5), 789-794. doi:10.3758/PP.70.5.789

de Fockert, J., & Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces. *The Quarterly Journal of Experimental Psychology*, *62*, 1716. doi:10.1080/17470210902811249

Haberman, J., & Whitney, D. (2011). Efficient summary statistical representation when change localization fails. *Psychonomic Bulletin & Review, 18,* 855-859. doi:10.3758/s13423-011-0125-6

Marchant, A., Simons, D., & de Fockert, J. (2013). Ensemble representations: Effects of set size and item heterogeneity on average size perception. *Acta Psychologica*, *142,* 245-250. doi:10.1016/j.actpsy.2012.11.002

Neumann, M., Schweinberger, S., & Burton, A. (2013). Viewers extract mean and individual identity from sets of famous faces. *Cognition*, *128,* 56-63. doi:10.1016/j.cognition.2013.03.006

Rensink, R. A. (2002). Change detection. *Annual Review of Psychology*, *53*, 245-277. doi:10.1146/annurev.psych.53.100901.135125

Simons, D., & Levin, D. (1997). Change blindness. *Trends in Cognitive Sciences*, *1,* 261- 267. doi:10.1016/S1364-6613(97)01080-2

Tottenham, N., Tanaka, J., Leon, A.C., McCarry, T., Nurse, M., Hare, T.A., Marcus, D.J., Westerlund, A., Casey, B.J., Nelson, C.A. (2009). The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry Research*, *168*, 242-249.

Appendix

*Figure 1*. Example stimulus

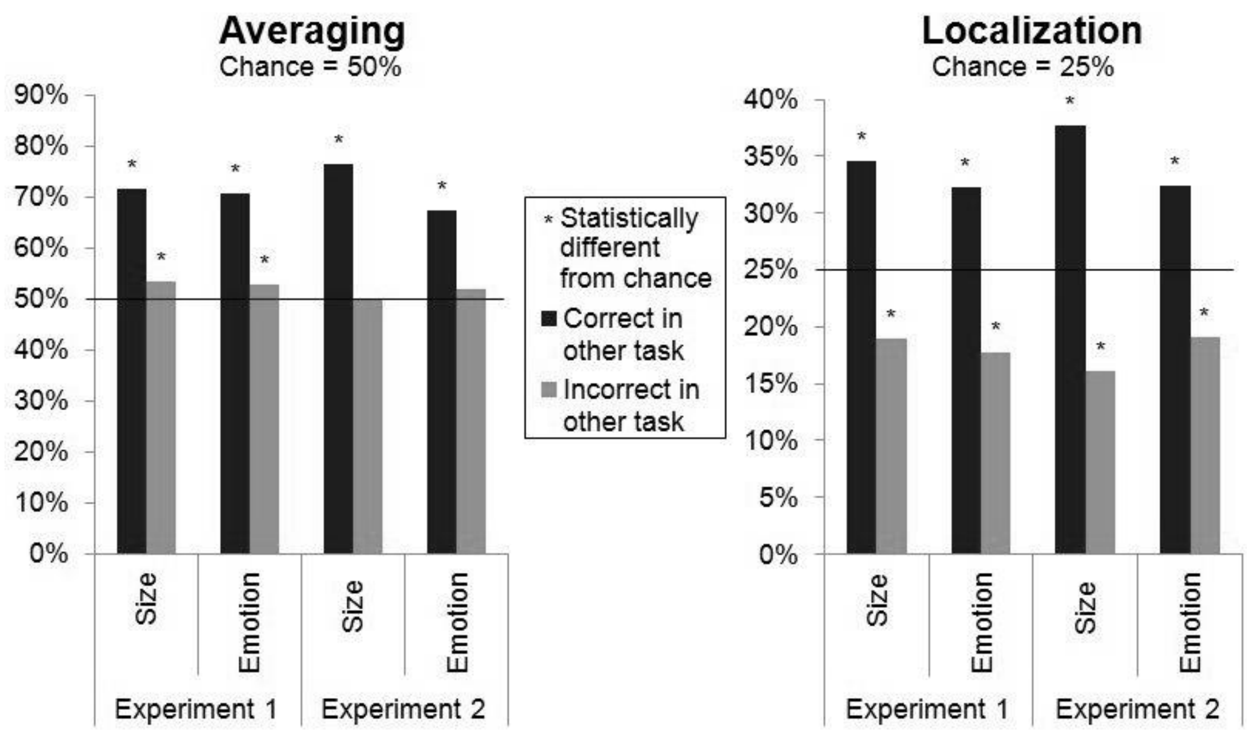


*Figure 2*. Performance on the mean discrimination and localization tasks categorized by success or failure on the other task.