Does Inversion Disrupt Averaging of Emotional Expressions?

An Honours Thesis
Submitted to the Department of Psychology
In Partial Fulfillment of the Requirements
For the Degree of

Bachelor of Arts (Honours)
in
Psychology
Luther College

by

Sarah Elyse Schimmel
Regina, Saskatchewan
April, 2018

Copyright 2019: Sarah Schimmel
Abstract

Previous research suggests subjects compute statistical summary representations (SSRs) to represent the average emotion of a set of faces without representing the individual faces comprising the set. Evidence for this claim relies on the finding that subjects can identify changes to average expression even when they cannot localize any face that changed. However, previous work in our lab suggests subjects must perceive changes to individual faces to infer changes in average expression. Thus, it is unclear whether subjects can compute average expression without encoding individual faces. Inverting faces impairs recognition of some individual facial expressions (sadness, disgust) more than others (fear, neutral). Subjects judged which of two consecutive arrays of upright or inverted faces displayed more sadness or disgust, or were more fearful or neutral. Finding that 1) inversion interferes with recognition of fear/neutral in ensembles of faces and 2) inversion has no effect on recognition of average sadness/disgust in ensembles would demonstrate a double dissociation between the processing of individual items and ensembles, and provide strong support for the claim that computing the average expression of a set of faces does not rely on individual exemplars. Results showed a double dissociation between the processing of individual items and ensembles. The inversion effect was statistically significant for items of disgust, but not for ensembles, and was not statistically significant for neutral items, but was for ensembles. The inversion effect was statistically significant for items and ensembles of both fear and sadness.
Acknowledgments

I would like to express my sincere gratitude to Dr. Chris Oriet for supervising my Honours thesis. Your commitment of support and instruction was of unparalleled effort. It was your profound understanding and passion for your craft that challenged me to surpass my limitations. Please know that your mentorship has changed my life, for which I am grateful.
Dedication

I would like to thank my friends and family for the incredible amount of support and encouragement they provided. I could not imagine the last few years without the tremendous support I received from loved ones. I especially want to thank my husband James for his unbelievable faith in me. I am thankful for the sacrifices he made without expectation nor hesitation. Especially when my confidence waned, he tirelessly rekindled my motivation to continue this work.

I also want to thank my Honours cohort, and congratulate them on their accomplishments. As some of the most kind, driven, and brilliant people I know, I am confident that they will each produce a legacy that will impact our community for the better. I am honoured to be considered part of this group.
Does Inversion Disrupt Averaging of Emotional Expressions?

Individual objects that fall within the focus of attention can often be described in rich detail. A single item, such as a wildflower, when carefully inspected can be full of colour, texture, and unique characteristics. However, when presented with a field full of wildflowers the observer perceives an abstract or “gist” representation of the field. This set representation discards local visual features and enables the observer to encode a larger amount of visual information and circumvent the limited capacity of visual working memory (Ariely, 2001).

According to Ariely (2001), when an observer is presented with an individual item, focused attention can be allocated to the location of the item and local features including characteristics such as its size and location can be encoded. In contrast, when an observer encounters groups of similar objects, such as a field of wildflowers, a general gist or average of the group is encoded without representing the local visual features of individual items.

Ariely (2001) found that when people see a group of similar objects the visual system creates a representation of those objects that is different than the representation of a single object in that group. Ariely suggested that a representation of a set of objects is a more efficient way for a limited capacity system to deal with a large amount of visual information. This suggests that sets of objects are represented differently than the individual objects that make up the set. Further building on Ariely’s research, Alvarez and Oliva (2008) found evidence that people are able to determine characteristics such as the average size or colour of groups of objects such as a bed of flowers, a shelf of books, or a crowd of people. Group representation, as described above, is often referred to as a statistical summary representation (SSR). SSRs, unlike individual object representations, are abstract and devoid of individual details (Alvarez & Oliva, 2008).
Researchers suggest that individual items and SSRs have different attentional requirements (Alvarez & Oliva, 2008). Individual items, which are observed in rich detail, require focused attention, whereas SSRs, or a group average of similar objects, require attention that is diffused across the set. Alvarez and Oliva found that the visual system compensates for information presented outside the focus of attention by summarizing it according to its statistical properties. Simply put, SSRs may allow people to represent information beyond the focus of attention (Haberman & Whitney, 2011). Unsurprisingly, research has shown that the representation of visual objects that fall outside the focus of attention does not capture the same details as attended objects (Alvarez & Oliva, 2008). Despite the limited information available for items that are observed without focused attention there does not appear to be complete blindness for these items. This could suggest that some information outside the focus of attention is more robust to the withdrawal of attention than other kinds of information (Alvarez & Oliva, 2008).

To better understand the impact of attention withdrawal Alvarez and Oliva (2008) ran three different experiments in which they manipulated the focus of attention by instructing participants to attend to target items at the expense of distractors. In their study, they showed that localizing a distractor was difficult and resulted in chance performance. However, subjects performed well when asked to locate the mean of the distractor locations. Alvarez and Oliva suggest that to achieve greater accuracy in localizing the distractor “centroid” location, or the mean location, subjects must be pooling noisy signals of the individual items in each location into a SSR.

**Emotion Expression Ensembles**

Haberman and Whitney (2011) tested whether individual items and SSRs are affected by change blindness, which occurs for items presented outside the focus of attention. In doing this,
the researchers showed subjects an array, or ensemble, of sixteen photos of the same face with variation in facial expressions. They presented two different arrays successively in which the four most expressive faces changed in expression from the first to the second array. Subjects were asked to perform two tasks. First, subjects had to identify a face that changed in the display. Second, subjects had to identify which array appeared, on average, happier. By asking subjects to identify a particular face that changed in the display, the researchers were able to determine whether attention was directed to a particular location in the display. If subjects were unable to localize a change in the array but were nevertheless able to determine the direction of change of the average expression, then it would suggest that SSRs can be computed without attending to individual items.

To address this question the researchers excluded from analysis trials on which localization was successful. Performance exceeded chance on trials on which localization was unsuccessful. Simply put, it appeared as though people were able to identify the overall change to the face ensemble even when they were unable to identify one of the items that contributed to this change. This suggests that people are able to compute SSRs even when they do not attend to individual items.

Haberman and Whitney asked subjects to both localize an individual face change in the set and to determine how the average expression of the set changed as a whole. If subjects were able to localize an individual face change in a set, it would be a simple task for subjects to guess the change in the average expression because the change in one item is perfectly correlated with the change to the set as a whole. However, subjects were still able to identify changes to the group with above chance performance on trials on which localization was unsuccessful. By asking subjects to localize individual face changes, Haberman and Whitney were essentially
asking subjects to focus attention on target objects. Research suggests that when observers focus on a particular object it makes diffusion of attention across the display more difficult (Oriet, Giesinger, & Stewart, 2018). Oriet et al. conducted an experiment similar to Haberman and Whitney’s but removed opportunities for subjects to infer changes to the group from changes to individual items. In their study, success of averaging appeared to be strongly correlated with success of localization. They found that when subjects were unable to localize, their ability to average was no greater than chance. Therefore, it is brought into question whether people are able to average emotional expression ensembles because performance could be accounted for by assuming subjects rely on changes to individual items in the set.

Furthermore, individual face recognition and identification are believed to be predominantly dependent on configural processing (Burton, Schweinberger, Jenkins, & Kaufmann, 2015). Some researchers suggest that configural processing likely requires focused attention (Eimer, Holmes, & McGlone, 2003) which may make diffusion of attention across the set more difficult. Burton et al. delineate three different types of configural processing. The first type is first-order relations, which refers to the general placement and configuration of facial features such as eyes, lips and nose within the face. The second type of configural processing describes a more holistic processing of the face. The third type is second-order relations, which refers to the spatial distances among features.

The Inversion Effect

The inversion effect, a phenomenon that disrupts face recognition, occurs when an image of a face is inverted (i.e., turned upside down). In most situations inversion has little effect on the recognition of the inverted item. However, inversion has a notable effect on an observer’s ability
to determine a face identity (Yin, 1969). The effect that inversion has on face recognition is distinct and research has been conducted to determine why faces are affected so uniquely.

As discussed earlier, face recognition is believed to rely on configural processing which researchers suggest is a large component of why face recognition is so vulnerable to the effect of inversion (Burton et al., 2015). Diamond and Carey (1986) suggest that for an item to be vulnerable to the inversion effect it must meet three criteria. First, members of the class must share a common configuration. Faces meet this criterion with first-order relations of configural processing. Second, observers must be able to individuate between members as a result of second-order relations. Again, faces meet this criterion because the spatial distancing between facial features is an important part of both recognizing and differentiating between identities. Second-order relations are a critical component of face recognition and inversion modifies the spatial layout of facial features (Burton et al., 2015). Third, observers must be experts in recognition of the class. This is also true for faces, because people have an astonishing expertise in face recognition (Diamond & Carey, 1986.). People see an incredible number of faces every day, whether at work, at school, or even in the media, and they must be able to individuate these faces despite the similarities.

Dernstl, Seidel, Kainz, and Carbon (2009) studied the effect of inversion on emotional expression recognition. They did this by comparing accuracy for recognition of upright or inverted images of five emotional expressions (anger, disgust, fear, happiness, and sadness) and a neutral expression. The authors found that inversion appears to disrupt recognition of positive and negative emotional expressions. The difficulty recognizing a distinct facial feature, such as a smile, suggests that emotion recognition is influenced by the configuration of features, and not just the features themselves. When inverted, some emotional expressions appear to be more
difficult to recognize than others. For example, sadness and disgust appear to be particularly disrupted by inversion. Again, the difficulty subjects have recognizing both inverted sad and disgusted faces suggests that recognition of these emotions is heavily dependent on configural processing.

Neutral expressions and emotional expressions of fear appeared to be the least affected by facial inversion (Derntl et al., 2009). Recognition of neutral expressions not only appeared to be unaffected by face inversion, but also little affected by presentation time. Accuracy was around 90% for neutral expression trials whether upright or inverted, or whether limited or unlimited in duration. In contrast, recognition of sadness and disgust appeared to be significantly affected by facial inversion. On trials with faces expressing sadness or disgust accuracy dropped dramatically with inversion, sometimes by more than 20%. Reducing exposure duration decreased accuracy of recognition of both disgust and sadness. The authors note that facial expressions of disgust vary more than other emotional expressions.

Taken together, the available evidence suggests that inversion had a significant effect on the recognition of emotional expressions. Emotional expressions involve specific configuration of facial features and inversion is believed to disrupt the normal spatial layout of those features (Burton et al., 2015). Moreover, recognition of emotional expressions appears to be heavily reliant on second-order relations (Burton et al., 2015). Recognition of some emotional expressions, like sadness and disgust, appear to experience a larger cost of inversion than others.

Purpose

The purpose of this study was to determine whether people are able to compute the average emotional expression from an array of faces. I tested two hypotheses: (1) if computing the average expression of an ensemble of faces relies on attention to individual exemplars, then
inversion will interfere with the ability to recognize emotional expressions from face ensembles in the same way that it interferes with individual items, and (2) if performance does not rely on attending to individual items in the set, inversion will interfere with the ability to recognize the emotional expression of individual items but will have no effect on determining the average expression. Thus, there were three possible predictions. (1) If people are unable to compute SSRs for emotional expressions without attending to individual exemplars then the effect inversion has on individual exemplars will be the same as the effect inversion has on averaging emotional expressions. (2) If SSRs do not depend on individual exemplars then recognition accuracy for inverted expressions of disgust and sadness will have similar accuracy as recognition of upright expressions. Put simply, averaging emotional expressions will not be disrupted by inversion like recognition of individual expressions is disrupted. It is also possible that inversion will affect all types of emotion recognition in ensembles even if it only affects sadness and disgust in individual exemplars. Thus, an alternative possibility that would be consistent with Hypothesis 2 is that (3) expressions of sadness, disgust, fear, and neutral expressions in sets of faces will all be disrupted by inversion in the same way. Both Outcome 2 and 3 are consistent with the inference that inversion affects recognition of emotional expressions in ensembles differently than it affects emotion recognition in individual faces.

**Method**

**Participants**

Despite the apparently large effect of inversion, a more conservative medium effect size was estimated. A power analysis completed using More Power (Campbell & Thompson, 2012) in which alpha was set at .05, with an effect size of a partial eta square of .06 determined a total
of 26 subjects would be required to achieve power = .80. The researcher chose to recruit approximately 40 subjects in the event that there was an error in data collection.

Thirty-seven students, 30 females and 7 males \((M = 21.95, SD = 5.13)\), from the University of Regina participated in this study. Subjects voluntarily signed up for the study through the University of Regina Participant Pool and were awarded one percent bonus credit in a course of their choosing as compensation for their time. In order to be eligible to participate in the study subjects were required to have normal or corrected-to-normal vision. Before beginning data collection subjects were provided with a written consent form. While subjects read and signed the consent form they were informed that their credit was being awarded to them through the Participant Pool website.

Due to incomplete data collection data from three subjects were removed from analysis. Data from one additional subject were removed because the subject performed at 50% accuracy, or no better than chance.

**Emotion Pretest**

In the emotion pretest subjects saw a single face on each trial that was paired with a label. Face images were taken from the NimStim Database (Tottenham et al., 2009) which were gray-scaled, cropped to remove any exterior features, and modified in size to be 100 x 120 pixels. Pretest images were identical to those used in the ensembles in the main task. Each trial displayed one of four emotional expressions (i.e., sadness, disgust, fear, or neutral) for 1000 ms. Half of the images were displayed in an upright orientation and the other half were displayed in an inverted orientation. The image was then removed and followed by a black screen for 500 ms. A label then appeared on the same black background and remained on the screen until the subject responded. The label provided was either a match or a mismatch to the previous emotion
expression shown according to the categorization reported by Tottenham et al. (2009). Subjects were asked to respond with one of two keys, pressing “1” for “same” and “2” for “different” on the number pad of the computer keyboard.

Before trials began the subject was supplied with written instructions for the task. The researcher carefully supervised the subject, gave oral instructions, and answered any questions before beginning practice trials. The subject completed 10 practice trials in which no feedback was provided. The researcher monitored all practice trials and once it was clear that the subject understood the task the researcher enabled the subject to continue on with the pretest trials. Subjects completed 160 pretest trials.

**Main Task**

**Stimuli.** Subjects were shown an array of 16 upright or inverted faces which were presented in a 4 x 4 grid pattern. Images were grey-scaled, cropped to remove exterior features, and displayed on a black background. The same gray-scaled face images were used in both the pretest and the main task. There was a total of 10 female and 10 male identities that were used in the study. Each trial consisted of either all female or all male identities. Each array displayed identities randomly among the 16 locations. Each identity chosen had one image available in the NimStim database for each emotional expression used (i.e., sadness, disgust, fear, and neutral) (Tottenham et al., 2009). Face images were morphed so that each identity had an image that was 50% disgust and 50% sadness and an image that was 50% fear and 50% neutral. Morphing emotional expressions allowed for varying degrees of emotional intensity in ensembles. All face images were face mapped and morphed using InterFace software (Kramer, Jenkins, & Burton, 2017).
Trials were grouped into two conditions, one condition consisted of sad and disgusted images and the other condition consisted of fearful and neutral images. Trials consisted of a mixture of 100% and 50% emotion expression images and were created to display predominantly one of the two emotions in the condition. The total amount of emotion expressed in an array was computed by summing the proportion of the emotion in each face (.5 or 1) and dividing this sum by the number of faces in the array (16). An example ensemble can be seen in Figure 1. In order for a randomly-generated array to be used in a trial it needed to meet the criterion that the total proportion of one of the two emotions did not exceed 35% (e.g., 65% disgust and 35% sadness).

**Procedure**

The general design and procedure of this experiment was based off of Haberman and Whitney’s (2011) study. On each trial the subject saw an array of faces, all of which were either upright or inverted, for 1000 ms, followed by a blank screen for 500 ms. Following this delay, a second array of faces, which matched the orientation of the first array, was displayed for 1000 ms. This was followed by an additional black screen that remained until the subject responded. This delay was terminated by a fixation cross displayed at the centre of the screen for 500 ms, which was immediately replaced by the first array when subjects pressed either the “1” or the “2” key. Subjects were asked to choose which of the two arrays, on average, appeared to display more of one emotional expression. Subjects were allowed an unlimited response time and were asked to respond as accurately as possible by choosing one of two keys that corresponded to the first or the second array.

Trials were grouped into four blocks. Two blocks had arrays comprised of a combination of sad and disgust images, and two blocks had arrays comprised of a combination of neutral and fearful images. Each block began by giving the subject instructions and telling them which
emotion expression to look for during that block of trials. For example, for one block of sad and disgusted images subjects would be asked, “which array, on average, is sadder?” whereas for the other block of sad and disgusted images subjects would instead be asked, “which array, on average, is more disgusted?” Each block had a total of 48 trials.

After completing the pretest trials and before subjects could begin the main task they were required to get the researcher. Subjects were given written instructions on how to complete the task while the researcher stayed to answer any questions and to clarify the task if needed. The researcher stayed to supervise while the subject completed 10 practice trials which provided them feedback. If the subject chose incorrectly they would get a red negative sign on the monitor; however, if they chose correctly they would get a green positive sign on the monitor. Once the practice trials were completed the researcher confirmed that the subject had a thorough understanding of the task. Subjects were then asked to complete the four blocks of trials. Once completed subjects were given a debriefing form and given an additional opportunity to ask questions.

**Results**

To test the effect of inversion on recognition of emotional expressions a 4 (Emotion Type: fear, neutral, disgust, and sadness) x 2 (Orientation: upright and inverted) repeated measures analysis of variance (ANOVA) was carried out. Three separate repeated measures ANOVAs were conducted to analyze three different dependent variables; the pretest response time, the pretest accuracy, and the main task accuracy. Paired samples t-tests were conducted to compare each emotion expression recognition rating with its inverted counterpart to index the cost of inversion.
**Pretest Response Time**

To examine the effect of inversion on emotion expression recognition a 4 x 2 repeated measures ANOVA was carried out. To accurately examine the effect of inversion on response time all incorrect trials were excluded from analysis. Initial data analysis revealed outliers so all trials that had a response time of 200 ms or less or 4000 ms or more were excluded from analysis. A 2.5 standard deviation cut off was employed, excluding 4.44% of the remaining data as outliers. The results show a statistically significant main effect for Emotion Type \( F(3, 108) = 33.63, \text{MSE} = 14756.83, p < .001, \eta_p^2 = .48 \) and a statistically significant main effect for Inversion \( F(1, 36) = 9.42, \text{MSE} = 12365.66, p = .004, \eta_p^2 = .207 \). There was no statistically significant interaction for Emotion Type and Inversion \( F(3, 108) = 1.30, \text{MSE} = 13458, p = .28, \eta_p^2 = .04 \).

On average, subjects were faster to recognize neutral and disgusted expressions than fearful or sad expressions. The cost of inversion was largest for sad expressions (83.86 ms), but was still evident for neutral expressions (33.01 ms), fearful expressions (29.36 ms) and disgusted expressions (12.45 ms). To further analyze the effect of inversion on the recognition of emotional expressions paired sample t-tests were carried out. The Bonferroni correction was used to correct for multiple comparisons, so all comparisons were tested with an alpha of .0125. For images of disgust the mean for upright trials was \( M = 862 (SD = 212.87) \); the mean for inverted trials was \( M = 875 (SD = 223.16) \). This difference between these means was not statistically significant, \( t(36) = .72, p = .48 \). For images of fear the mean for upright trials was \( M = 962 (SD = 177.69) \); the mean for inverted trials was \( M = 991 (SD = 254.34) \). This difference between these means was not statistically significant, \( t(36) = 1.01, p = .32 \). For images of neutral expressions the mean for upright trials was \( M = 770 (SD = 176.09) \); the mean for inverted trials
was $M = 803$ ($SD = 209.22$). This difference between these means was not statistically significant, $t(36) = 1.36, p = .18$. For images of sadness the mean for upright trials was $M = 885$ ($SD = 208.88$); the mean for inverted trials was $M = 969$ ($SD = 207.82$). This difference between the means was statistically significant, $t(36) = 2.52, p < .016$.

**Pretest Accuracy**

To further test the effect of inversion on emotional expression recognition response accuracy was analyzed. A 4 x 2 repeated measures ANOVA was employed to examine recognition accuracy. Pretest Accuracy scores violated Mauchly’s Test of Sphericity, so degrees of freedom were adjusted using the Greenhouse-Geisser correction. Results showed a statistically significant main effect for Emotion Type, $F(2.28, 81.98) = 25.95, \text{MSE} = .02, p < .001, \eta_p^2 = .42$, a statistically significant main effect for Inversion, $F(1, 36) = 80.74, \text{MSE} = .01, p < .001, \eta_p^2 = .69$, and a statistically significant interaction for Emotion Type and Inversion, $F(2.58, 92.76) = 24.34, \text{MSE} = .009, p < .001, \eta_p^2 = .40$.

On average, subjects performed best on emotion recognition on upright trials with neutral and sad expressions, but sad expressions showed the most severe cost of inversion. There was a significant interaction between Emotion Type and Inversion in pretest accuracy, so paired sample t-tests were conducted. The Bonferroni correction was used to correct for multiple comparisons, so all comparisons were tested with an alpha of .0125. In contrast to response time, inversion had a strong effect on the accuracy of recognition for all emotional expressions except for neutral. For images of disgust the mean for upright trials was $M = .89$ ($SD = .10$); the mean for inverted trials was $M = .82$ ($SD = .10$). This difference between the means was statistically significant, $t(36) = 3.5, p < .001$. For images of fear the mean for upright trials was $M = .82$ ($SD = .14$); the mean for inverted trials was $M = .74$ ($SD = .15$). This difference between the means
was statistically significant, $t(36) = 4.28$, $p < .001$. For images of sadness the mean for upright trials was $M = .90$ ($SD = .08$); the mean for inverted trials was $M = .64$ ($SD = .15$). This difference between the means was statistically significant, $t(36) = 8.68$, $p < .001$. For images of neutral expressions the mean for upright trials was $M = .93$ ($SD = .095$); the mean for inverted trials was $M = .91$ ($SD = .09$). This difference between the means was not statistically significant, $t(36) = 1.73$, $p < .09$.

**Main Task Accuracy**

To examine the effect of inversion on the accuracy of emotion recognition in face ensembles a 4 x 2 repeated measures ANOVA was conducted. The results showed a statistically significant main effect for Emotion Type $F(3, 108) = 24.69$, $MSE = .009$, $p < .001$, $\eta^2_p = .41$, a statistically significant main effect for Inversion $F(1, 36) = 11.21$, $MSE = .02$, $p = .002$, $\eta^2_p = .24$, and a statistically significant interaction for Emotion Type and Inversion $F(3, 108) = 5.298$, $MSE = .008$, $p = .002$, $\eta^2_p = .13$.

On average, subjects performed best on upright trials when judging neutral and disgusted expressions, but sad, neutral, and fearful judgment trials showed the biggest cost of inversion. To further analyze the interaction between Emotion Type and Inversion paired sample t-tests were conducted. The Bonferroni correction was used to correct for multiple comparisons, so all comparisons were tested with an alpha of .0125. For ensembles of disgust the mean for upright trials was $M = .70$ ($SD = .08$); the mean for inverted trials was $M = .71$ ($SD = .10$). This difference between the means was not statistically significant, $t(36) = .73$, $p = .47$. For ensembles of fear the mean for upright trials was $M = .74$ ($SD = .09$); the mean for inverted trials was $M = .67$ ($SD = .12$). This difference between the means was statistically significant, $t(36) = 3.10$, $p = .004$. For ensembles of neutral expression the mean for upright trials was $M = .80$ ($SD = .08$); the
mean for inverted trials was $M = .73$ ($SD = .11$). This difference between the means was statistically significant, $t(36) = 2.75, p = .009$. For ensembles of sadness the mean for upright trials was $M = .68$ ($SD = .12$); the mean for inverted trials was $M = .58$ ($SD = .17$). This difference between the means was statistically significant, $t(36) = 3.22, p = .003$.

**Discussion**

I hypothesized that if inversion disrupts the recognition of emotion expression ensembles in the same way that it disrupts the recognition of individual emotion expressions, then SSRs must rely on individual exemplars; however, if the pattern of results for emotion expression ensembles is different than the pattern of results for individual expressions, then performance does not rely on individual exemplars. The results suggests that for at least some expressions, people are able to average emotional expressions without attending to a particular item in the set. Thus, the current study supports the results found by Haberman and Whitney in which subjects were able to average second-order relations across a set.

Although the effect of inversion on pretest response time was not significant, the results are in the direction expected if inversion disrupts emotion recognition. To better understand the effect of inversion on ensembles, it was important to examine the cost of inversion on recognition accuracy at the individual level. There was both a main effect for emotion type and inversion, as well as an interaction between emotion type and inversion for pretest accuracy. Further testing showed that the effect of inversion on accuracy was significant for sadness, disgust, and fear, but was not significant for neutral trials. These results somewhat differed from Derntl et al.’s findings which showed that although the cost of inversion on accuracy was significant for expressions of fear, the cost of inversion on fear was dramatically smaller than on expressions of disgust and sadness. In contrast, the current study shows the cost of inversion was
larger for expressions of fear than disgust. In the study by Derntl et al. the researchers mentioned five previous studies that examined the effect of inversion on emotional expression recognition. Out of the five studies mentioned, four of the studies showed a significant effect of inversion on the recognition of expressions of fear (Fallshore & Bartholow, 2003; Goren & Wilson, 2006; McKelvie, 1995; Prkachin, 2003).

To test the effect of inversion on emotional expression recognition a two-alternative forced choice matching task was carried out. This method of testing the effect of inversion was a novel design that had not been used in previous studies. Subjects were shown an individual face that was followed by an emotional expression label in which subjects had to decide if the label was a match or a mismatch as quickly as possible. In contrast, Derntl et al. provided subjects with an individual face and a list of six possible emotional expression labels on the right hand side of the screen. Perhaps the simplicity of the matching design in the current study could have influenced the effect of inversion on the subject’s performance of emotional expression recognition.

There was a main effect for both emotion type and orientation, as well as an interaction between emotion type and orientation for main task accuracy. Further testing was done to examine the effect of inversion on face ensemble recognition which showed that, except for disgust, inversion had a statistically significant cost to the recognition of emotion expression ensembles. Ensembles of disgust were not influenced by the inversion effect unlike expressions of fear, sadness, and neutral expressions. These unique results suggest that some emotion expression types are not affected by inversion when processed as ensembles. This thought-provoking discovery provides more insight into processing emotional expression ensembles and will hopefully encourage further investigation into this novel finding.
Of considerable importance is the effect that inversion had on neutral expressions and expressions of disgust in the pretest and the main task. Inversion had a significant cost to the recognition of individual expressions of disgust, but did not have a significant cost to averaging ensembles of disgust. In contrast, inversion did not have a significant cost to the recognition of individual neutral expressions, but did have a significant cost to averaging neutral ensembles. The disparity between the cost of inversion on neutral and disgusted expressions during the pretest and the main task strongly suggests that averaging ensembles is a different process than recognizing individual emotional expressions. These results show a notable double dissociation between the processing of individual expressions and emotional ensembles. The results of this study strongly suggest that processing ensembles may require a process that is unique from that of individual recognition. Thus, the results of the present study inform the current debate over the role of focused attention on ensemble processing (e.g., Cohen, Dennett, and Kanwisher, 2016).

As discussed earlier face recognition is believed to rely on second-order relations which are thought to require focused attention (Eimer, Holmes, & McGlone, 2003). If this were the case then subjects would rely on attending to individual items in the set to accurately average the face ensemble. However, the results of the current study suggest that subjects are able to compute the average emotional expression of face ensembles without attending to individual items in the set, thus suggesting that subjects are able to aggregate second-order relations in configural processing without focused attention.

Trials in the main task were grouped into four blocks in which each block had participants attend to one of the four emotional expressions. However, there were only two block types, either a combination of sad and disgusted faces or a combination of fearful and neutral
faces. Therefore, subjects saw the same arrays for both the block of trials in which they attended to sadness and the block of trials in which they attended to disgust. Results show that the cost of inversion was significant for the block of sad trials, but was not significant for the block of disgust trials. Therefore, the significant difference in results can only be explained by the difference in what the subject was asked to attend to. This suggests a “top-down” process that is guided by the instructions given to the subject.

**Limitations and Future Directions**

The current study took place at the University of Regina and subjects signed up through Department of Psychology Participant Pool to volunteer in this study. A majority of subjects were female and likely had some post-secondary education; therefore, it is possible that subjects that participated in this study are not representative of the population.

Although the current study was based off of Haberman and Whitney’s (2011) study there were some notable differences in stimuli and general design. Many of the changes made to the design of the current study were intended to prevent subjects from focusing on individual items and generalizing changes in individual items to changes to the set as a whole. By preventing the subjects from focusing on a particular face and location that changed between sets the researcher aimed to determine if ensemble performance was due to individual items. It is possible that some of those changes could have resulted in limitations to the current study. In the study conducted by Haberman and Whitney there was only one identity used in each array, whereas in the current study there were a total of 10 female and 10 male identities. It is possible that in the current study subjects may have found some identities visually stood out among the others. Although this may have occurred it is unlikely that it altered the findings of the study. Both the placement of the
identity and the degree of emotion intensity was randomized making it difficult to track individual changes between the sets.

Another possible limitation to the study was the significant cost of inversion to the recognition of fear in the pretest. Ensembles were created to pair two expressions that were strongly affected by inversion at the individual level and two that were not. During the pretest trials the cost of inversion was significant for recognition of expressions of fear. This was not the expected finding and, although unlikely, may mean that something about the current stimuli set of fearful and neutral emotion expressions was unusual. Future studies should examine this more closely and try to disentangle this possible variable. In the future, different emotion combinations and emotional expressions should be examined in order to better understand the possible impact of grouping emotional expressions.

Although unlikely, it cannot be ruled out that subjects were able to quickly scan a small percentage of the array and infer the average from that. However, if this had occurred the results still suggest that face ensembles are processed differently than individual faces. If recognition of a small proportion of faces used the same process as individual emotion expression recognition then the results from the averaging task would have replicated those found in the pretest. Nevertheless, future studies should further examine the possibility that performance for averaging emotional expression ensembles is due to attending to individual items in the set. As suggested by Oriet et al. (2019) subject performance with face ensembles could be tested with a gaze contingent eye tracker and a central focus point. If subjects attend to a particular face for a prolonged period of time the image would be removed from the screen. The use of a gaze contingent eye tracker would ensure that participants are unable to focus on an individual item in the set and provide researchers with more valuable insight into emotion expression averaging.
Conclusion

Results from this study suggest that people are able to average emotion expression ensembles, which supports some previous studies. Notable findings suggest that emotional expression ensembles do not rely on attending to individual items in the set, and that second-order relations can be aggregated across an ensemble. Furthermore, novel findings from the current study show that some emotional expression ensembles do not appear to be affected by inversion. This study adds to the literature about attention and perception with respect to how people perceive large amounts of stimuli when focused attention is unavailable.
References

https://doi.org/10.1111/j.1467-9280.2008.02098.x


http://doi.org/10.1016/j.tics.2016.03.006


https://doi.org/10.1037/0096-3445.115.2.107

https://doi.org/10.2466/pms.2003.96.1.236


https://doi.org/10.3758/s13423-011-0125-6


https://doi.org/10.1348/000712603762842093

Appendices

Figure 1. Face Ensembles
Actual trials consisted of all female or all male ensembles. There were a total of 10 female and 10 male identities used in the study. Only certain images from the study are permitted for publishing under the license agreement for accessing the database the images were taken from, and therefore the ensembles shown are meant to serve as examples of what subjects saw. The image on the left is an upright ensemble of fear and neutral expressions and the image on the right is an inverted ensemble of sad and disgusted expressions. Images are from the NimStim Database (Tottenham et al., 2009).
Figure 2. Pretest Response Time and Accuracy.
The numbers in the brackets represent corresponding accuracy scores for pretest trials. Pretest trials consisted of individual face images that were followed by a label. All pretest trials were a forced choice response where subjects indicated whether the label was a match or a mismatch to the previous emotion expression. Highlighted accuracy scores denote statistical significance for the cost of inversion, where * $p < .05$ and ** $p < .01$. 
Figure 3. Main Task Accuracy.
Main task trials consisted of two face ensembles that were displayed in sequential order. The dashed line represents 50%, or chance, accuracy. Highlighted columns denote statistical significance for the cost of inversion, where * $p < .05$ and ** $p < .01$. 