Mood Judgment: Can People Accurately Judge the Mood of a Crowd?

Jessica Kardasz

Rider University

Abstract

When looking at a crowd of people, how accurate are we at judging the overall mood when we are only able to focus on a portion of faces? Through 24 participants, the current study aimed to understand how accurate we judge moods of a crowd, if people who are more socially perceptive having a higher accuracy of estimation, and how directional changes impact final judgments. A 3 (Starting Emotion: negative, neutral, positive) x 3 (Ending Emotion: negative, neutral, positive) within-subjects factorial design was used. Participants completed the Reading the Mind in the Eyes Task (RMET) and viewed 9 flicker trials which had directional changes in emotion. We found that the emotion of the first slide has a significant impact on final judgments. We found a large correlation between RMET scores and mean error. Replication and further research are needed to understand the findings presented in this study.

Keywords; emotion, social perceptiveness, crowd judgment, flicker

Mood Judgment: Can People Accurately Judge the Average Mood of a Crowd?

At different points in our lives, we are placed in positions which require us to scan a crowd and make decisions on what the overall mood is. However, within a crowd, emotions change rapidly and can jump from a positive mood to a negative one within a matter of seconds. With this, how are we to be sure that what we believe about the overall mood of a group is accurate? It is impossible to look at all faces within a group at once, which results in us having to look at a portion of faces that may or may not define the mood of the rest. Understanding the emotions of others is important because it allows us to appropriately navigate and respond to our surroundings, as well as better communication with those around us. While there may be some level of difficulty in fully understanding how a crowd feels, is there a group of people who are better at judging emotions than others? In the current study, we will be taking a deeper look into how accurately people judge the emotion of a crowd, and if people with higher social perceptiveness perform better in these situations.

For the current study, we were greatly influenced by the concept and results discussed in Goldenberg, Weisz, Sweeny, Cikara, and Gross (2021). Here, they aimed to understand how people read the emotion of the crowd and how they make the decision on what the overall mood is. They hypothesized that when ranking the mood of a crowd, that they will overestimate when more intense emotions are present, which they refer to as the crowd-emotion-amplification effect. Additionally, they believed that this effect would be much greater with the presence of negative emotions as well as when there are more faces in the array. Across 3 separate experiments, participants were exposed to an array of 1-to-12 faces varying in intensity of emotion, viewing a condition of either neutral-to-happy or neutral-to-angry. In both experiments 1 and 2, people did overestimate the mood of the group, even when there was an increase in

exposure time. For experiment 3, they implemented an eye tracking measuring which showed that people are no more likely to look at more emotional faces, but when they do, they look at these faces for a longer period of time which may increase this amplification and that there was no difference between negative and positive faces. While Goldenberg et al. produced some interesting findings, a key limitation they had was that throughout all experiments and trials, they used the face of only one individual, which is not realistic and makes it harder for results to generalize. In the current study, we aimed to avoid this by actively trying to limit having duplicate faces on the same slide as well as having a set of 28 facial stimuli, all varying in emotion. Additionally, similar to how there was a neutral-to-happy and neutral-to-angry condition, we implemented more conditions in which overtime throughout the trial, emotion would rapidly change.

Throughout the years, previous research has looked into the concept of judging the emotion of a crowd and the accuracy of doing so. In Haberman and Whitney (2009), they looked into the idea of ensemble coding in terms of emotions. Ensemble coding, as they discuss, is about how when we have these complex things in our environment, we process only the gist and fill in the gaps. Haberman and Whitney hypothesized that participants in a task where they are told to focus on an individual within a set of faces, that they will do better in judging the overall mean emotion of the crowd. Essentially, how accurate are these judgments? For this study, they went through several experiments where they would expand on their original hypothesis and tasks as well as replicate procedures. Throughout the experiments, participants were exposed to a set of 4 to 16 faces which varied in emotion at a given time and were asked to make judgments about the mood of the group. They found that people are generally good at estimating the overall mood and that people are capable of processing the emotion of large groups. With the findings

and methods of both Goldenberg et al. (2021) and Haberman and Whitney, we decided that if we were to present participants with quick glances at a crowd and implement directional changes throughout trials, that we could grasp a greater understanding of how accurate humans are in judging a crowd.

While research has shown that people are generally good at estimating the mood of a crowd, and that they may overestimate at times, is there a group of people who can do this more effectively? When we make quick decisions, about overall mood, what gives us the most information? The findings in Leach, Ammar, England, and Remigio (2016) suggest that sometimes having less information leads to more and that this information can be gathered solely through the eyes. In this study, they aimed to understand how people can detect liars versus truth-tellers through niqabs, a type of veil that covers most of the face, with only the eyes being shown. Here, they set up a mock theft where participants watched a video of a woman either watching a bag or also stealing from it and then they were asked to testify on behalf of that woman, being randomly assigned to either a hijab, niqab, or no veil. Leach et al. hypothesized that lie detection would be more accurate in the niqab condition. They found that participants were much more accurate in detecting whether or not the witness was a liar or a truth-teller when they wore niqabs, compared to those with no veil. This illustrates that the eyes can play a major role in how we understand others and that sometimes more information can result in less accurate outcomes. With these findings, we were able to better understand that in the way that less information is more, that quick glances at varied emotions may be more beneficial in estimating the overall mood. Additionally, this led us in the direction that high social perceptiveness may be a potential group that could better estimate the emotion of a group.

This idea of social perceptiveness has been brought about and tested through a task called the Reading the Mind in the Eyes Test (RMET), which aims to understand how well people can identify the emotion of others. Baron-Cohen, Jolliffee, Mortimore, and Robertson (1997) looked into testing theory of mind, specifically in individuals with Autism and Asperger's Syndrome. Theory of mind deals with understanding mental states in both others and the self and Baron Cohen et al. tested this in high-functioning adults by showing them photographs of faces of varying emotions, but only with the eyes showing. Here, participants were required to decide on which mental state, out of 2 options, was the best fit. The results suggest that there is a difference between one's general intelligence and one's level of social perceptiveness. For the current study, we used a digitized version of the RMET to understand if people who are more socially perceptive, meaning those who are capable of guessing emotion accurately, since they may be better when they are only provided with quick glances of emotions. Being socially perceptive is important in understanding our interactions with others. By picking up these cues more easily, it is possible that this group of people can grasp a more accurate understanding of the overall mood of a group.

In the present study, we examined whether or not people can accurately judge the average emotion of a group. Here, we hypothesized that similar to previous research, people will generally be able to estimate emotion and that people may overestimate in the presence of very negative or very positive emotions. Next, while previous research has shown that people can provide general estimates, what happens when there are rapid directional changes of emotion and how does this change over time impact one’s final judgment? We hypothesized that the emotion in which participants are originally exposed to in the beginning of the trial will negatively impact their ability in making an accurate final judgment. Lastly, are people who are more socially

perceptive going to be more accurate in their judgments of a crowd? For our final hypothesis, we predicted that people who score higher on the RMET will judge emotion more accurately, despite directional changes.

Method

Participants

Twenty-four relatives and friends of the experimenters were recruited as the participants for this study, resulting in a power of .67. All participants were over the age of 18 and provided informed written consent. The current study was reviewed and approved by the institutional review board at Rider University.

Materials

For this study, all participants were exposed to the same materials and completed tasks on both paper and through a laptop. The core materials, the facial stimuli were emojis that demonstrated the variable moods of negative, neutral, and positive. There was a total of 28 emojis that were yellow with varying facial expressions, which we gathered through an extensive Google Search. With these emojis, we performed a manipulation check which illustrated that these emojis measured the emotions which we wanted them to measure, and these results are shown in both Figures 1 and 2.

For the Faces task, participants viewed 9 trials, with each having emotion go in a different direction. These directions included: no change positive, no change negative, no change neutral, negative to positive, negative to neutral, positive to negative, positive to neutral, neutral to positive, and neutral to negative. For each trial, 16 faces were randomly placed at some point on a 6x6 grid. For negative slides, there were a total of 6 neutral and 10 negative faces, with 5 being very negative and 5 slightly negative. With the neutral slides, there were 5 slightly

positive, 5 slightly negative, and 6 neutral faces. For the positive slides, there were 6 neutral and 10 positive faces, with 5 slightly positive and 5 very positive. An important aspect of the Faces task is that each trial flickered anywhere from 2 to 10 times. The exposure time for each slide with faces was 1 second, with flicker time being .25 seconds. Following each flicker, every face on the slides changed. These trials included flickers to make participants unable to predict the end point of emotion and to take away any ability of strategizing on what the mood of the final slide would be. Following each trial, a slide came up asking them to rank the average emotion of the crowd.

To rank the emotion, participants were provided with a double-sided response sheet which included bars for a practice trial and the 9 experimental trials. These bars ranged from 0, which represented a very negative mood overall to 165, being very positive. The instructions on the response sheet asked participants to draw a line on the scale to indicate where they believed the overall mood of the group was. We measured these lines in millimeters with a ruler.

Apparatus

For the RMET, participants used a MacBook Pro 2017 to complete the task. Using the same device, participants viewed all 9 trials through a Microsoft PowerPoint slideshow presentation.

Procedure

On the day of the experimental session, participants walked into an empty room and sat at a table with the experimenter sitting on the opposite side. Once settled, participants were prompted to provide their written informed consent. Following this, experimenters directed participants to take the RMET which requires participants to guess the emotion of 37 individuals, with 1 practice trial and 4 emotion options for each question (Hagen, n.d.). Additionally, the

scale for the RMET ranged from 0 to 36, with 36 representing high social perceptiveness. Once completed, participants were then presented with the Faces task which consisted of 9 flicker trials and 16 faces on each slide. The variable number of flickers in each trial ranged from 2 to 10. Prior to completing the 9 experimental trials, participants were given a static practice trial to understand what they were to look for when ranking emotion on the response sheet. Following each experimental trial, participants were asked to rate the general mood of the crowd in the last slide. Participants marked a line on the response sheet to where they felt the general mood of the crowd was. The laptop for this task was placed away from participants at an arm's length where they could click the arrow button to move on to the next slide to limit experimenter bias. Once all trials were completed, participants were informally debriefed, where they were made aware of what the study was specifically measuring. The entire procedure lasted approximately 20 minutes.

Design

The current study was a 3 (Starting Emotion: negative, neutral, positive) x 3 (Ending Emotion: negative, neutral, positive) within-subjects factorial design. For this study, we had a total of 4 dependent variables. The first related to mood rating and was based on how far up the line participants marked in millimeters from the inner edge of the far-left side to the inner edge of the left side of the line they made. Next, another dependent variable was participant's social perceptiveness scores which were gathered from the RMET task. Furthermore, we also had a dependent variable for total absolute estimation error as well as average mean estimation error. For controls, all subjects were exposed to all conditions and they all viewed the trials in the same randomized order. For randomization, we utilized Research Randomizer to determine the

placement of faces on slides, the number of flickers for each trial, and the order in which trials were presented to participants.

Analysis

To analyze the data, we conducted 2 3x3 within-subjects factorial ANOVAs through VassarStats. With ANOVA, we compared the mood ratings across the 9 trials to the start and end mood, as well as the interaction between the two. Additionally, we also used ANOVA to compare the accuracy of estimation, based on if the faces were supposed to represent a negative, neutral, or positive mood overall, and to see the interaction between start and end mood. For this study, we had a Type II error rate of $\beta = .33$, a Type I error rate of $\alpha=.05$, and a d_{MEI} of .50. We also performed 2 Pearson's r to correlate RMET results with mean and total absolute error of how participants did on the Faces task, with a d_{MEI} of .25.

Results

A summary of the results for mean slide ratings is provided in Figure 3. The main effect of start mood was significant, $F(2,46) = 39.86$, $p < .001$, $\eta = .43$. The main effect of end mood was significant, $F(2,46) = 36.37$, $p < .001$, $\eta = .43$. The interaction between start and end mood was significant, $F(4,92) = 2.97$, $p = .023$, $\eta = .186$.

Figure 4 provides a summary of results for accuracy of estimation. The main effect of start mood was significant, $F(2,46) = 31.21$, $p < .001$, $\eta = .42$. The main effect of end mood was not significant, $F(2,46) = 6.11$, $p = .004$, $\eta = 1.95$. The interaction between start and end mood was significant, $F(4,92) = 4.08$, $p = .004$, $\eta = .24$.

The results of the correlation between RMET results and the accuracy of estimation are presented in Figure 5. The total absolute error of RMET and the accuracy of estimation was not

significant, r = -.296 [CI: -.562, .235]. The mean error of RMET and the accuracy of estimation was significant, r = .440 [CI: -.721, -.035].

Discussion

The present study examined how accurately we are able to judge the emotion of a crowd and how directional changes may impact final judgments. Additionally, we aimed to understand if people who received higher social perceptiveness scores on the RMET had a greater accuracy of estimation than those with lower scores. Previous studies have shown that generally, humans are relatively good at estimating the emotion of an array of faces, but that at times, overestimation can occur (Goldenberg, Weisz, Sweeny, Cikara, & Gross (2021); Haberman and Whitney, 2009). Furthermore, studies such as that of Leach, Ammar, England, and Remigio (2016) illustrated the concept of how sometimes less information can result in more information and that people do well at detection, even when this is solely through the eyes. With the findings of previous research, and the objective of the RMET, which tests one's ability to read emotion through the eyes, we tested whether people could accurately judge emotion through quick glances at an array of faces through 9 trials with directional changes and if people scores on RMET correlation with accuracy.

With this study, we had both significant and non-significant results. Starting off with mood ratings, we found that there was a significant difference in starting mood on end mood, as well as in the interaction between the two. Essentially, the mood that was present at the beginning of the trial mattered more than the mood at the end. This confirms our hypothesis that with directional changes, the emotion that is present at the beginning of the trial negatively impacts final judgments. These results imply that in situations where we must judge emotion, we

may potentially miss out on some changes and be slightly inaccurate about what the overall emotion is as time goes on.

Looking at accuracy of estimation, we wanted to understand if participants would overestimate emotion and how accurate their rankings were with the mean emotion of the slide. We found that there was a significant difference between the slides that ended negative and those that ended positive. For those that ended negative, participants tended to overestimate, which is similar to the findings of experiments 1 and 2 in Goldenberg, Weisz, Sweeny, Cikara, & Gross (2021). On the other hand, for slides that ended positive, participants tended to underestimate emotion. With these findings, there is a possibility that we are potentially more drawn to negative faces, due to the consequences that could come about when we experience these emotions in real-life. While there seems to be the overestimation of negative faces, it is important that replication is done to see if this effect fades away, similar to how it did in Goldenberg et al. There is no difference in the interaction between negative and positive emotions. However, with the neutral emotion, there was an interaction present.

With the correlation between RMET scores and the accuracy of estimation, we found no significant differences with participant's total absolute error on the Faces task. While the correlation was right on the border, it was not enough to make a difference. On the other hand, our results revealed that there was a negative correlation between RMET and mean error. These results illustrate that with higher scores on the RMET, there was a negative error in which people ranked faces as less positive than they actually were. With lower scores on the RMET, there was a positive error where people ranked faces as more positive than they actually were. Additionally, the correlation here appears to be large but, this requires much speculation due to

the potential of these findings being a fluke. This only makes the need greater for replication to see if this effect still holds up.

While there were some interesting results across this study, there were some significant limitations which are important to address in order for replication and future research purposes. First, one of our greatest issues is that we had a small sample size of 24, which gave us low power. If we meant .80 power, there is a chance that the findings presented here could be much different. Another issue regarding our participants is that all of our participants were people we knew. With this, the current study has no external validity. Future studies should aim to meet .80 power and to avoid using a convenience sample in order for results to generalize. Another limitation that was present in this study is our facial stimuli. Our facial stimuli were yellow-faced emojis, ones that people often text, add to posts, and even occasionally use as a joke. The problem here is that the way that we perceive these faces may be viewed completely different by others, which could have potentially altered results. To avoid this, future studies can work on either creating their own facial stimuli to represent emotions, gathering a variety of people to demonstrate emotions, or using different types of emojis.

The current study illustrated that people can generally judge the emotion of a crowd even when there may be changes in direction. However, the starting emotion, whether that is negative, neutral, or a positive one, has a significant impact on one's final judgment of emotion. Additionally, the emotion present in the beginning tends to be more important than that of the end. These findings also illustrated there is slight overestimation when the ending slides are negative. We found that there was a large negative correlation between RMET scores and mean error rates in the Faces task, which are interesting results that push the need for replication. The

results of this study are similar to that of previous research, but future research is needed to determine if these effects stand over time.

References

Goldenberg, A., Weisz, E., Sweeny, T. D., Cikara, M., & Gross, J. J. (2021). The crowd-emotion-amplification effect. *Psychological Science*, *32*(3), 437–450. https://doi-org.rider.idm.oclc.org/10.1177/0956797620970561

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. https://doi.org/10.1037/a0013899

Leach, A. M., Ammar, N., England, D. N., Remigio, L. M., Kleinberg, B., & Verschuere, B. J. (2016). Less is more? Detecting lies in veiled witnesses. *Law and human behavior*, *40*(4), 401–410. https://doi.org/10.1037/lhb0000189

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of child psychology and psychiatry, and allied disciplines*, *42*(2), 241–251.

Urbaniak, G. C., & Plous, S. (2013). Research Randomizer (Version 4.0) [Computer software]. Retrieved on April 28, 2021, from http://www.randomizer.org/

Figure 1.

Manipulation check of starting mood (negative, neutral, positive) on mean face rating (negative, neutral, positive).

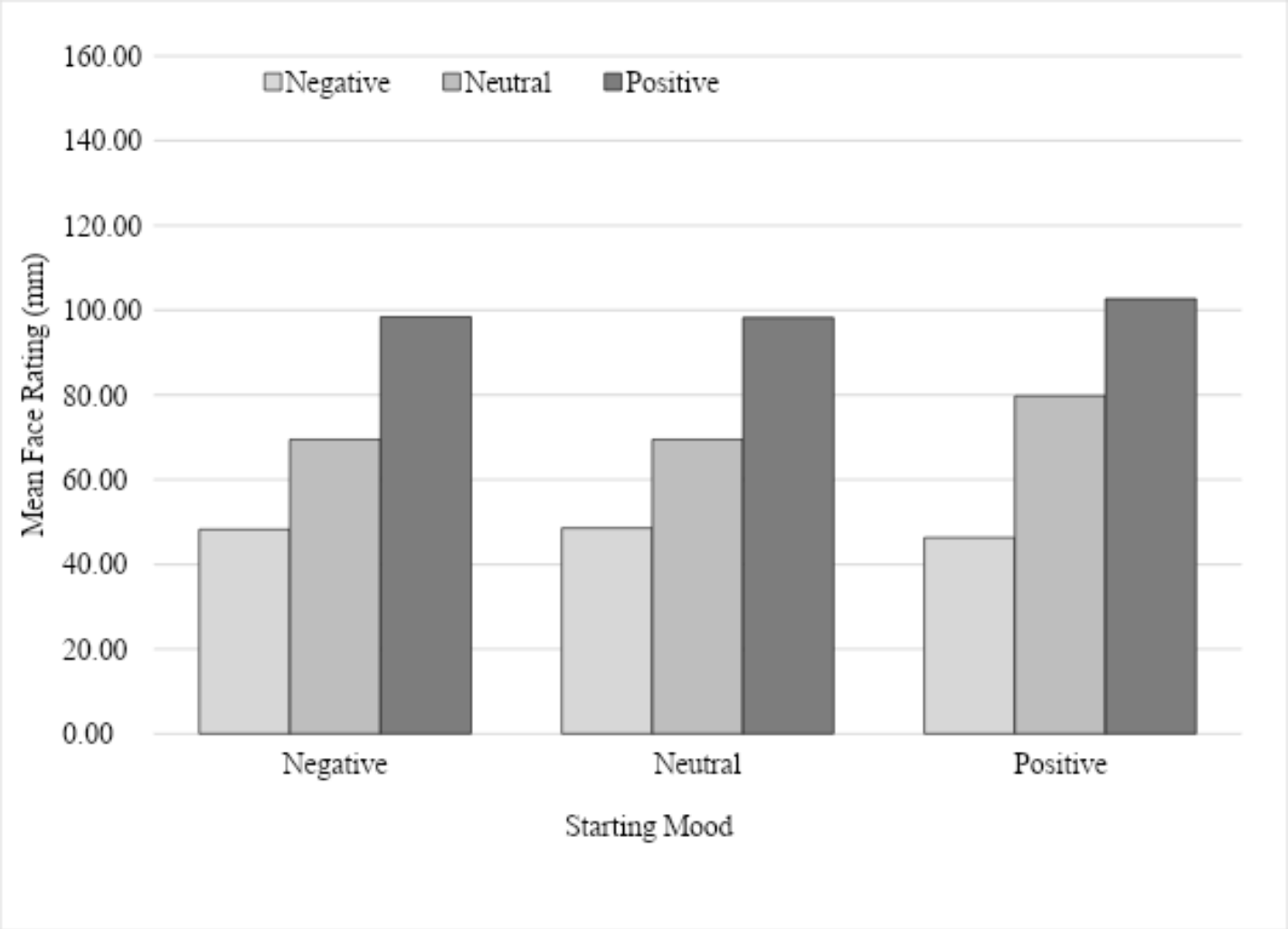

Figure 2.

The 28 facial stimuli ranked from lowest mean to highest mean, compared against mean mood ratings.

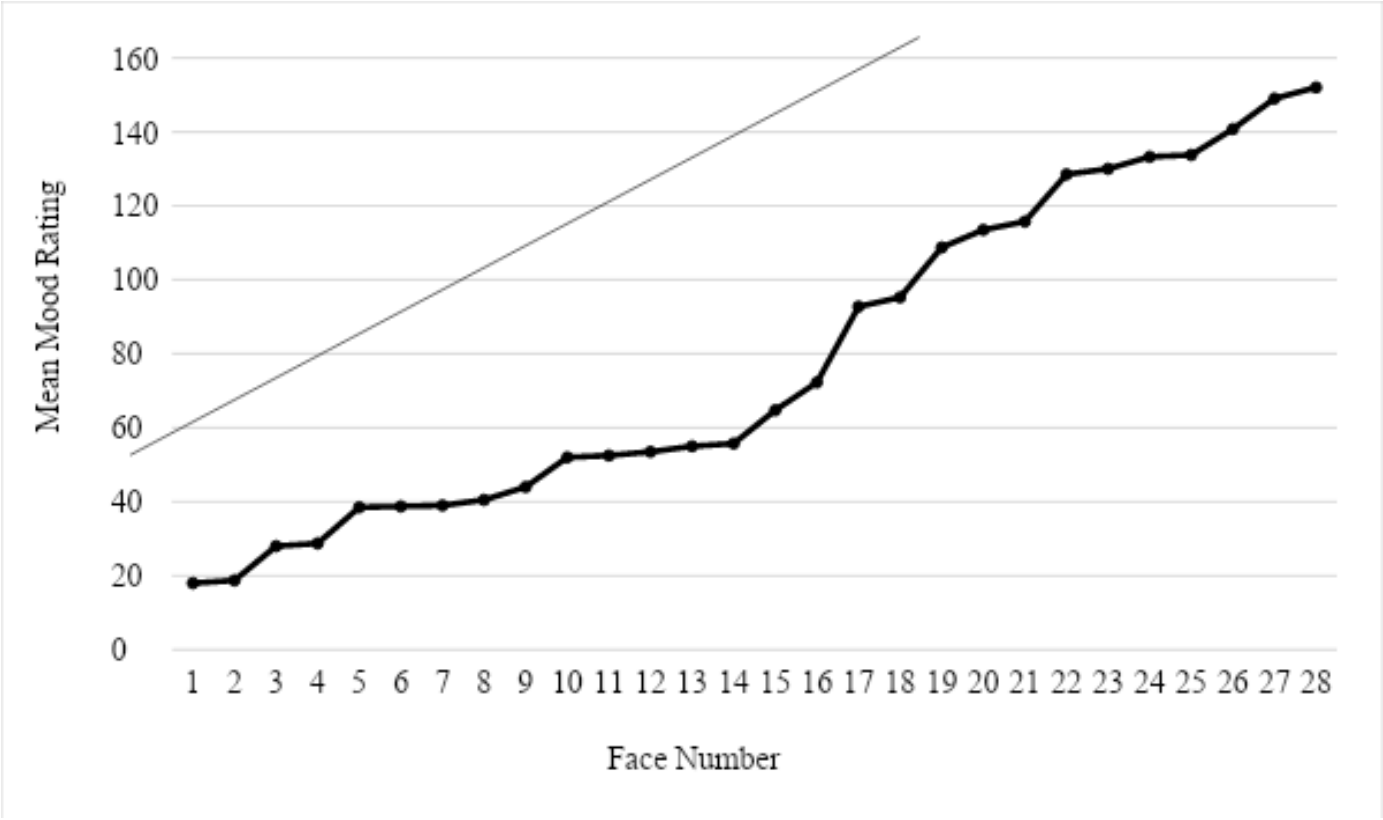

Figure 3

Mean mood ratings across all trials. Error bars represent (95%) confidence intervals.

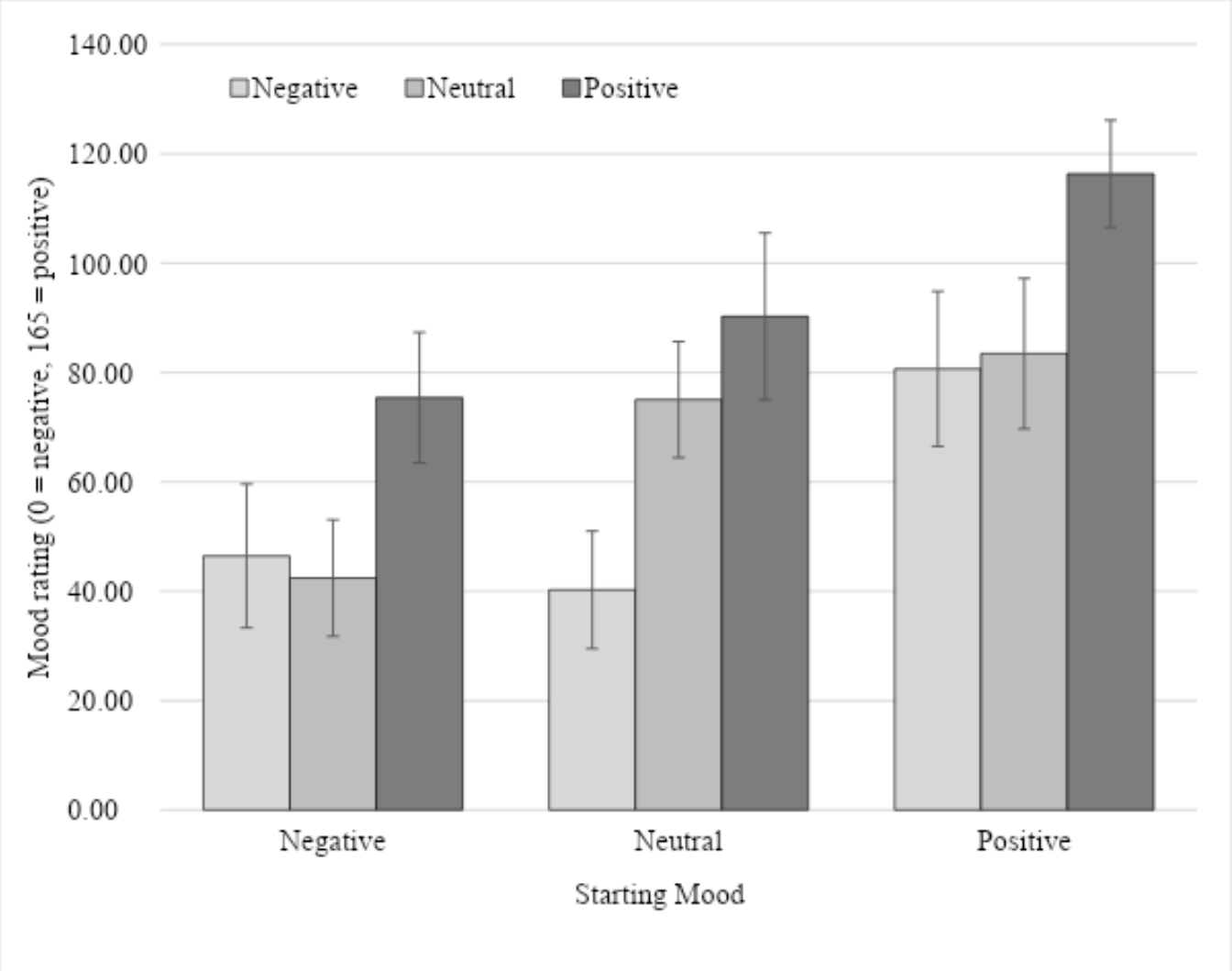

Figure 4

The accuracy of mood estimation based on what faces were supposed to represent (negative, neutral, positive). Error bars represent (95%) confidence intervals.

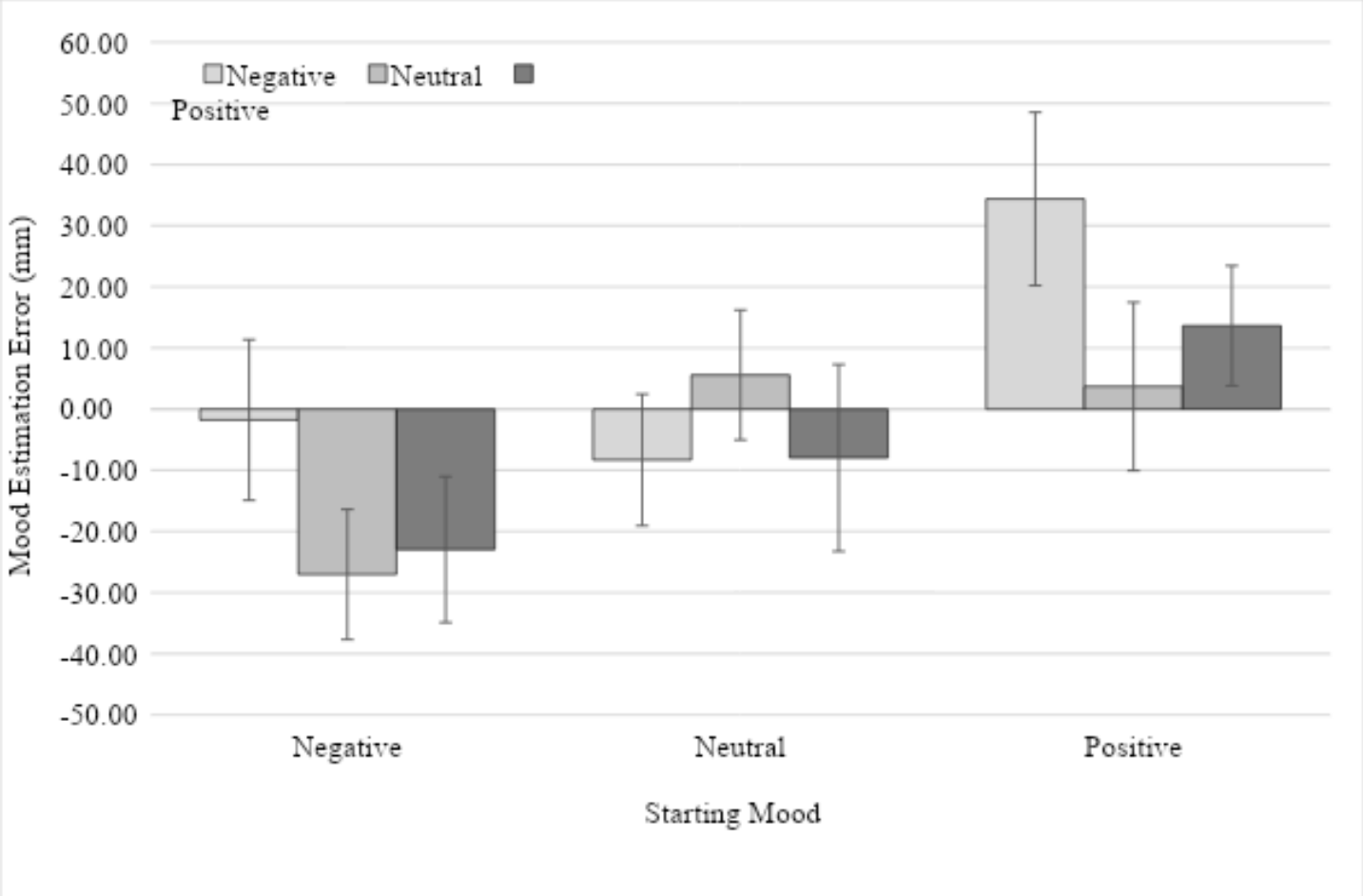

Figure 5

The mean error rates of mood ratings correlated with RMET scores.