STEMorph: Morphed Emotional Face Stimuli

Mohammad Ebrahim Katebi[1,2*], Mohammad Hossein Ghafari[3*], and Tara Ghafari[4]

[1] School of Cognitive Sciences, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran

[2] School of Medicine, Tehran University of Medical Sciences, Tehran, Iran

[3] School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran

[4] Centre for Human Brain Health, School of Psychology, University of Birmingham, UK

Author Note

[*] These authors contributed equally.

Correspondence concerning this article should be addressed to Tara Ghafari, Centre for Human Brain Health, School of Psychology, University of Birmingham, Birmingham, UK, B15 2TT. Email: t.ghafari@bham.ac.uk

Abstract

Emotion recognition is crucial for interpreting social cues, with facial expressions being the primary channel for such communication. Despite its importance, emotion recognition is often influenced by biases, in which we show a systematic recognition advantage for a particular emotion. These biases, however, are inconsistently reported across studies, likely due to methodological variations, underlining the necessity for a standardized approach. Traditional face morphing methods, although widely used, can create unnatural-looking stimuli, which may confound the interpretation of emotions. Addressing this issue, we here introduce *STEMorph*, a validated stimulus set based on the *NimStim* facial expression set. Our approach utilizes neutral-anchored morphing and neural-network-generated masks to ensure the natural appearance and integrity of the depicted emotions. we validated our stimulus set by presenting morphed emotional faces to participants and asking them to rate the emotional valence of each stimulus. The *STEMorph's* validity was confirmed through linear regression analysis, showing a strong correlation between subjective ratings and targeted emotional states. Additionally, subgroup analysis by gender of both the depicted faces and the participants showed uniform results. Moreover, we confirmed the reliability of *STEMorph* by asking the same participants to rate the stimuli two weeks later. In conclusion, by introducing a controlled, validated, and ecologically valid stimulus set of emotional faces, our study paves the way for further investigations aimed at unraveling the complexities of facial emotion recognition and deepening our understanding of this vital aspect of human interaction.

Keywords: *STEMorph*, Stimulus Set, Facial Emotional Expressions, Morphed Stimuli, *Nimstim*

Introduction

Emotion recognition is a fundamental aspect of human social interaction, enabling individuals to interpret and respond appropriately to the emotional states of others. Among the means through which emotions are conveyed, facial expressions stand as a primary modality, offering rich cues about an individual's internal states (Russell, 1994). However, despite its significance, the process of emotion recognition is not devoid of biases. These biases are systematic recognition advantages for particular emotions and have significant implications for both clinical and non-clinical populations (Silvia et al., 2006).

However, studies often report inconsistent results regarding the direction and magnitude of these biases (Münkler et al., 2015). For instance, one study found that healthy individuals did not demonstrate a bias toward happiness or sadness when presented with neutral facial expressions (Leppänen et al., 2004), while another study showed that healthy participants tended to perceive neutral faces as expressing happiness, whereas depressed individuals interpreted them as neutral (Surguladze et al., 2004). It is uncertain whether these mixed findings reflect genuine individual differences in emotion perception or are due to methodological variations between the studies (Vikhanova et al., 2022). This highlights the need for standardized approaches.

One commonly employed method involves using face morphing techniques. Face morphing is an image processing technique employed to achieve seamless transformation from one face to another and facilitates a gradual transition between two facial images. By morphing facial features between two extreme emotional expressions. researchers generate a spectrum of face images with ambiguous emotional expressions for participants to interpret (Joormann & Gotlib, 2006; Paiva-Silva et al., 2016). However, traditional face morphing presents challenges, particularly concerning the formation of unrealistic and unnatural facial stimuli (Paiva-Silva et al., 2016) which may confound interpretations and obscure real emotion recognition processes. Moreover, as various facial regions hold varying levels of importance for different expressions, the observed bias could reflect individual

and cultural divergences in participants' focus on specific facial areas or features (Elfenbein & Ambady, 2002; Schurgin et al., 2014; Vikhanova et al., 2022).

To maintain naturalism, proper masking of the face images (extracting the face area while removing the background) is required, as morphing introduces artifacts in non-face regions. Using no mask (full head) or simple oval/square masks which are employed by most studies (Dawel et al., 2022) may cause distortions that risk emotional content loss since various facial features hold different degrees of significance depending on the type of expression (Elfenbein & Ambady, 2002; Schurgin et al., 2014). However, there are no existing publicly available stimulus set without these methodological confounds (Dawel et al., 2022; Paiva-Silva et al., 2016; Testerman, 2023). To address this gap, we developed and validated a new stimulus set utilizing neutral-anchored morphing alongside neural-network-generated masks, to optimize naturalism and coherence.

We derived the stimuli from the *NimStim set of facial expressions (Tottenham et al., 2009)* which is a comprehensive collection of facial images portraying a spectrum of emotions including happy, angry, and neutral expressions. We adopt a nuanced approach to morphing emotional faces by generating distinct naturalistic face masks for each image using neural networks. Furthermore, we utilized the neutral facial expression as a central anchor point to enhance the naturalistic quality of the stimuli. We then evaluated the validity and reliability of *STEMorph* by asking a group of participants to provide ratings of the emotional content of these stimuli.

Method

Participants

A total of 27 participants (age range: 18-25, mean age: 20.85, 12 females) were recruited. The participants had no history of neurological or psychiatric disorder. Informed consent was obtained from all participants. This study was conducted with ethical approval from the ethics committee of Shahid Beheshti University of Medical Sciences.

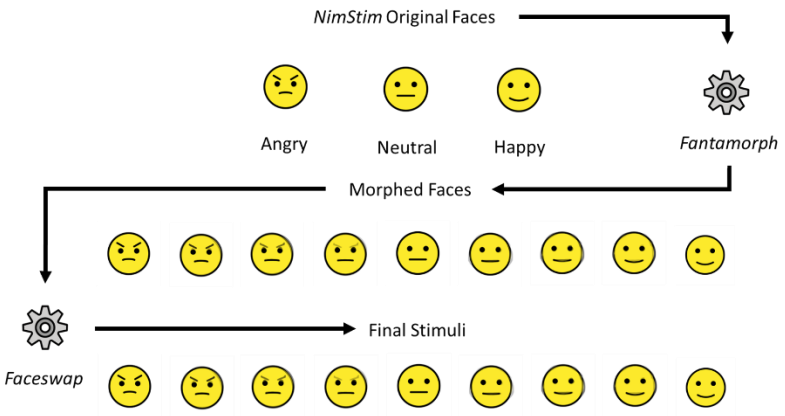
Experiment design & Face Generation

We used face morphing to produce a linear transition of emotions from anger to happiness using neutral expression as an intermediary point. The stimuli were derived from the *NimStim set of facial expressions (Tottenham et al., 2009)*. Each emotional expression has an open-mouth and closed-mouth variation in *NimStim*, from which we selected closed-mouth variations. *Abrosoft Fantamorph (Abrosoft Fantamorph Deluxe 5.4.8)* software was used for morphing, involving nine steps to transition from the angry face to neutrality and subsequently to the happy face. We used the original facial expressions of anger, happiness, and neutral as reference points and generated three intermediate emotional states between each original face.

We selected 22 individuals (11 males and 11 females) from *NimStim*, prioritizing those whose facial expressions facilitated smooth and natural morphing while avoiding individuals with extreme facial expressions (highly exaggerated or intense displays of emotions). The final set is composed of 198 distinct faces, representing 9 different emotional steps for each of the 22 individuals (Figure 1).

Figure 1

Example Stimuli from The Set.



Note. The top row illustrates the start (angry) and end (happy) points with intermediary morphing stages. The bottom row shows the corresponding neural-network-generated masks applied to each morphing step.

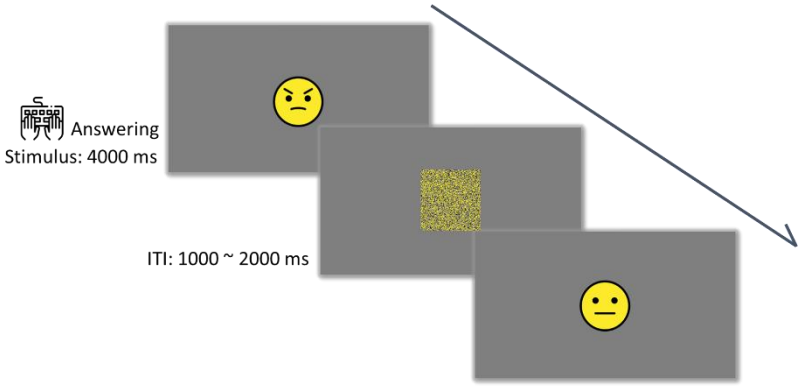
To address artifacts introduced in the ears, hair, and other non-landmark regions of the face during conventional face morphing, we generated distinct naturalistic face masks for each image using the *FaceSwap (Faceswap)* software, which utilizes the *Bisenet-FP (face-parsing.PyTorch)* neural network. *BiSeNet-FP* is a modified lightweight neural network designed for face parsing. The model was set to not include hair or ears and only the extracted face region. Manual editing was performed to correct for any artifacts arising from stray hair (in which a strand of hair overlapping the face resulted in a split in the face) or other extraneous features.

Experimental Setup

Participants were asked to rate the emotional state of each presented stimulus on a Likert scale from angry to happy. The task was developed and implemented using *Psychtoolbox (Brainard, 1997)* within *MATLAB R2022a* (Figure 2). Participants were comfortably seated in a dimly lit room facing a laptop monitor. The resolution of monitor screen was 1920 × 1080 and it was positioned ~54 cm from the participant. The stimuli were presented in a randomized order, with facial images appearing at the center of the screen. Participants were instructed to freely observe the images without fixation requirements.

Figure 2

Schematic representation of the experimental trial design.



Note. Participants were required to rate the emotional state of the face stimuli within a 4000 ms viewing window, followed by a randomized ITI of 1000-2000 ms.

The task started with pressing a key, after which participants were asked to rate the emotional state of each presented face based on a 9-point Likert scale. For half of the participants, 1 indicated the angriest emotional state and 9 indicated the happiest emotional state. For the other half, the 9-point scale was reversed. Each participant had a 4-second window to provide their rating; in the absence of a response within this timeframe, the task automatically proceeded to the subsequent face.

In total, participants completed 396 trials, interspersed by two 2-minute breaks. Following each trial, a 1500-ms inter-trial interval (ITI) ensued, during which a noise mask was displayed at the center of the screen to ensure no effect from interference of trials and priming (Aguado et al., 2015; Sweeny et al., 2009). To generate the noise mask the stimulus presented in the previous trial was scrambled into 4-pixel squares. Performance measures included reaction time and emotional state ratings. Reaction time was calculated as the time from the beginning of the stimulus presentation to the participant's response.

Reliability Assessment

To assess reliability, the same behavioral task was readministered to a subgroup of 16 participants two weeks following their initial participation.

Data Analysis

Validity

To investigate the relationship between participants' subjective ratings and the emotional morph level of the faces, we utilized the *scikit-learn (Pedregosa et al., 2011) Python* library to perform Ordinary Least Squares (OLS) linear regression. We calculated the coefficient of determination (R^2) to quantify the strength and direction of this relationship. To compare the calculated coefficient with the null hypothesis ($R^2 = 0$), we computed the Bayes Factor using the *BayesFactor (Morey & Rouder, 2023) R* package. Additionally, we determined the statistical power ($1 - \beta$ error) of the calculated coefficient

compared to the non-correlation model (correlation ρ H0 = 0) using *G*Power (Faul et al., 2009)* software with an α error of 0.05.

Reliability

To investigate reliability, we analyzed data from participants who engaged in the retesting (n = 16) and compared it with their validity data. We matched trials from the two tasks based on the presented face and performed OLS linear regression between participants' ratings in the validity task and the retest conducted two weeks later. We then calculated the Bayes Factor and Statistical power using the methods previously described.

Gender Subgroup Analysis

Data from the validity testing sessions were split according to the gender of the faces and subjects. OLS linear regression and Calculation of the coefficient of determination (R^2) were conducted for each group, as previously described.

Results

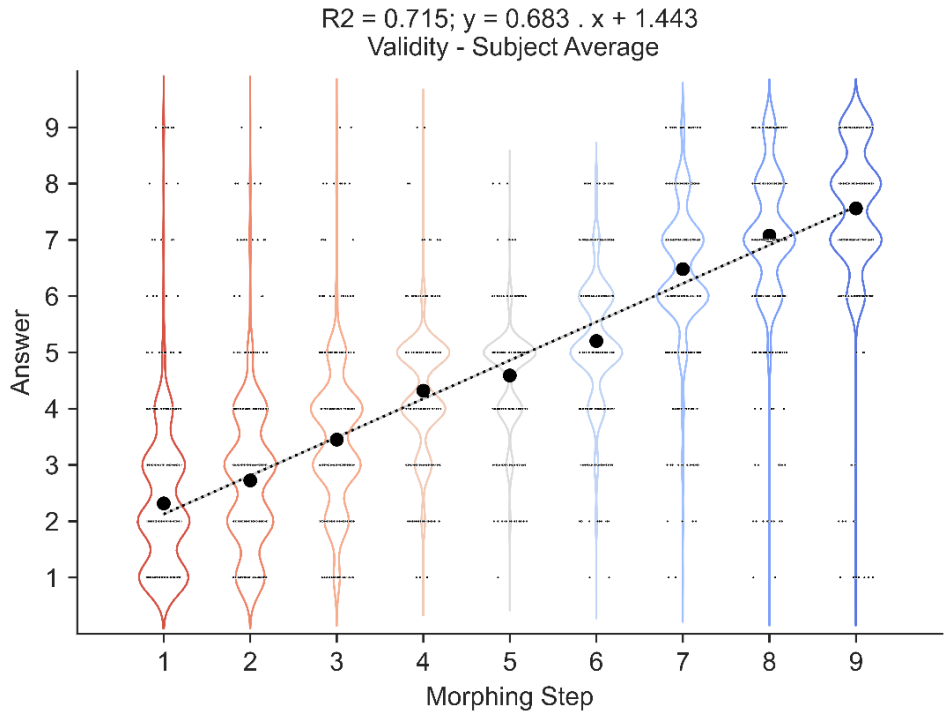
The aim of this study was to develop and validate a new stimulus set for emotion recognition utilizing neutral-anchored face morphing and neural-network-generated masks to optimize naturalism and coherence. A total of 27 participants were recruited to rate the emotional content of the face stimuli on a 9-point Likert scale ranging from angry to happy with a subset of 16 participants returning for a reliability assessment two weeks later.

Validity

We employed linear regression to examine the emotional ratings provided by participants (Figure 3). The morph step yielded a coefficient estimate of 0.683, with an intercept of 1.443. The calculated coefficient of determination (R^2) was 0.715 (p-value < 0.001). The Bayes factor to compare the calculated regression to the null hypothesis ($R^2 = 0$) was infinite. Furthermore, the calculated statistical power (1 − β error) in comparison with the non-correlation model (correlation ρ H0 = 0) was 0.999.

Figure 3

Validity of emotional ratings across morphing steps.



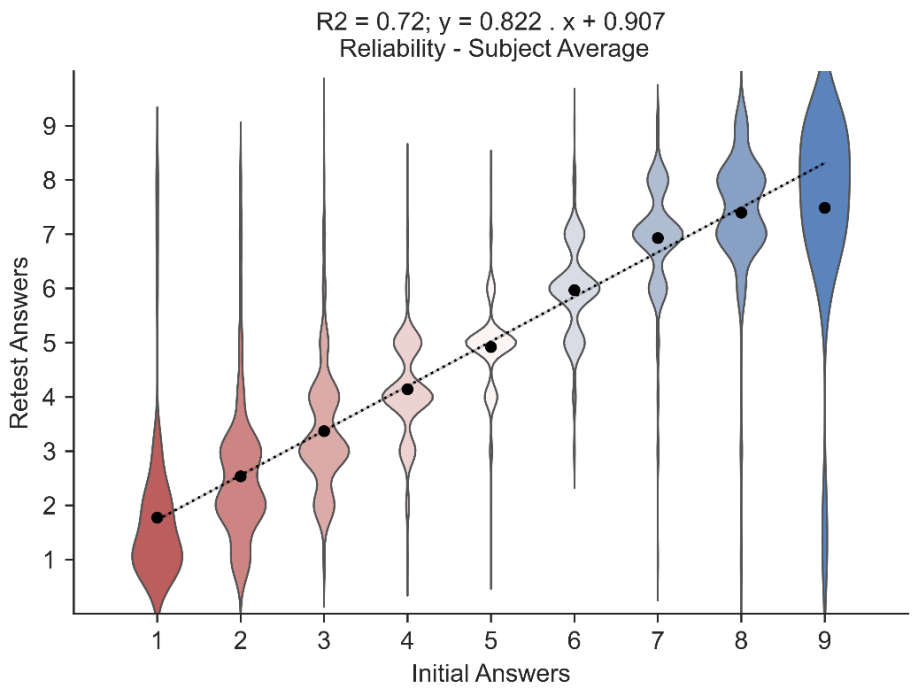
Note. The x-axis represents the morphing steps from angry (1) to happy (9), while the y-axis indicates the averaged participants' ratings of emotional content. The violin plots depict the distribution and probability density of ratings at each step, with individual trials represented by jittered black dots. The means are denoted by larger black dots, and the dashed line indicates the regression line. The regression equation and the coefficient of determination (R^2) are provided. The colors utilized are arbitrary and solely for presentation purposes.

Reliability

To evaluate reliability, we calculated the coefficient of determination (R^2) between the validity task and retest ratings, resulting in 0.72 (p-value < 0.001) (Figure 4). The coefficient estimate for the validity answers was 0.822, with an intercept of 0.907. The Bayes factor, comparing the calculated regression to the null hypothesis ($R^2 = 0$), was infinite. Moreover, the calculated statistical power ($1 - \beta$ error) in comparison with the non-correlation model (Correlation ρ H0 = 0) was 0.999.

Figure 4

Reliability of participants' ratings between the initial test and the retest session.



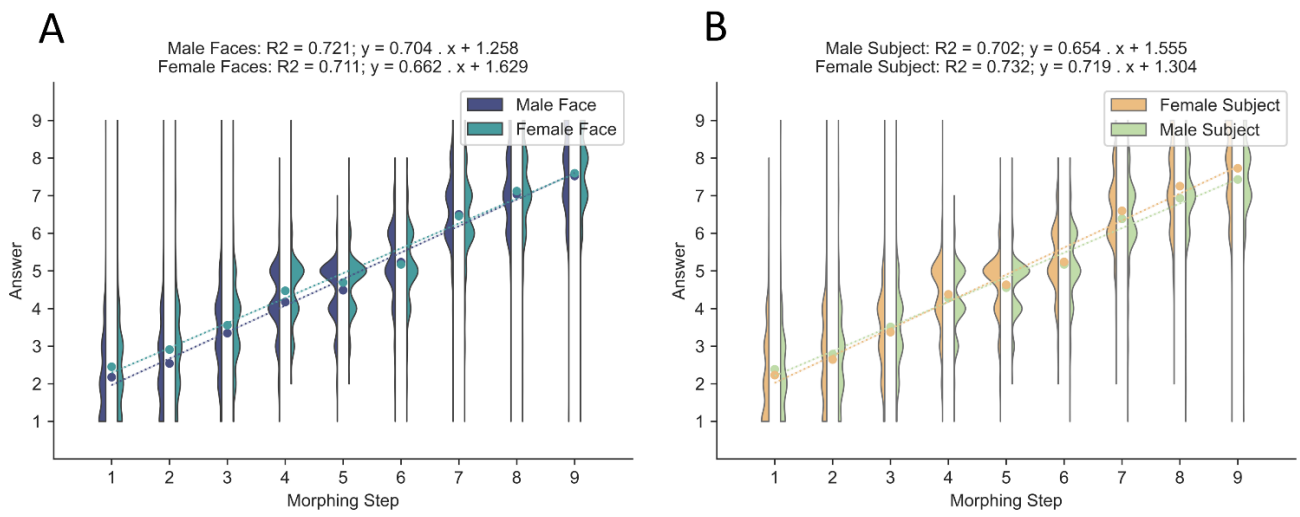
Note. The x-axis denotes the ratings from the first task, and the y-axis corresponds to ratings from the retest. Violin plots illustrate the distribution of ratings for each answer, with their probability densities indicated by plot width. The means of the retest ratings are marked with black dots, and the regression line is represented by a dashed line. The title includes the regression outcome and R^2 value. Colors are used arbitrarily for presentation purposes.

Gender Subgroup Validity Analysis

We conducted subgroup analysis for the gender of the faces in the stimulus set and the gender of the subject on the validity data, obtaining an R^2 value of 0.708 for male faces and 0.694 for female faces (See Figure 5.A). Furthermore, we calculated R^2 for male subjects, which was 0.702 and 0.698 for female subjects. (See Figure 5.B)

Figure 5

Gender-based subgroup analysis of the validity data.



Note. Panel A compares the validity of emotional recognition across morphing steps for male and female faces, while Panel B compares the validity ratings from male and female participants. Both panels include violin plots of distribution and a regression line.

Discussion

In this paper, we introduced a newly developed stimulus set of angry to happy facial expressions derived from the *NimStim* set, using neutral-anchored morphing and neural-network-generated masking. The findings provide strong support for the validity and reliability of *STEMorph*. Participants' subjective ratings showed a robust positive correlation with the intended emotional states. The high strength of this correlation demonstrates that our morphing methodology created a naturalistic continuum of emotional transformations. The high test-retest reliability suggests minimal variability in interpretations of this stimulus set across multiple exposures, indicating STEMorph provides a robust measurement tool for investigating emotion perception biases.

STEMorph is the first such contribution to the *NimStim* set, and we have employed an innovative method to create it. Using a neutral expression as an intermediary anchor point allowed us

to generate a spectrum of emotional faces transitioning smoothly between angry and happy, with minimum bias at the neutral point. Careful attention to masking techniques mitigated potential distortions, preserving the integrity of the facial stimuli. Masking of the face images (extracting only the face region and removing the background) is required to maintain a naturalistic face dataset, as the morphing procedure focuses on the face landmarks and introduces artifacts in the ears, hair, and other regions. We generated distinct naturalistic face masks using neural networks. Our analysis demonstrates that the STEMorph is both valid and reliable for gender-based subgroups, including female versus male faces and female versus male subjects.

Overall, by demonstrating reliability and validity with this initial stimulus set, our study sets the stage for expanded research to elucidate facial emotion recognition intricacies using controlled, validated, and ecologically sound stimuli. Public contribution of this set facilitates broader adoption, enabling more investigations that can further knowledge in this critical domain of human interaction. The strong linear relationships found provide a promising foundation. However, further examination across diverse samples could offer additional perspective into the generalizability and external validity of the patterns observed here. As the current stimulus set focused only on the anger-happiness continuum, developing complementary transitions between other emotions also offers potential for future investigation.

Declarations

Funding

No funding was received for conducting this study.

Conflicts of interest

The authors have no competing interests to declare that are relevant to the content of this article.

Ethics approval

Approval was obtained from the ethics committee of Shahid Beheshti University of Medical Sciences (IR.SBMU.MSP.REC.1396.757). The procedures used in this study adhere to the tenets of the Declaration of Helsinki.

Consent to participate

Informed consent was obtained from all individual participants included in the study.

Availability of data and materials

Code availability

The source code for the psychophysics task and data analysis performed in this study can be found in the following GitHub repository: https://github.com/Mekatebi/STEMorph

References

Abrosoft Fantamorph Deluxe 5.4.8. https://www.fantamorph.com/index.html

Aguado, L., Pedraza, I. S., & Gutiérrez, A. G. (2015). A comparison of backward masking of faces in expression and gender identification. *Psicologica*, *35*, 171-194.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spat Vis*, *10*(4), 433-436. https://www.ncbi.nlm.nih.gov/pubmed/9176952

Dawel, A., Miller, E. J., Horsburgh, A., & Ford, P. (2022). A systematic survey of face stimuli used in psychological research 2000–2020. *Behavior research methods*, *54*(4), 1889-1901. https://doi.org/10.3758/s13428-021-01705-3

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychological bulletin*, *128*(2), 203.

face-parsing.PyTorch. https://github.com/zllrunning/face-parsing.PyTorch

Faceswap. https://faceswap.dev/

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical Power Analyses Using G*Power 3.1: Tests for Correlation and Regression Analyses. *Behavior research methods*, *41*, 1149-1160. https://doi.org/10.3758/BRM.41.4.1149

Joormann, J., & Gotlib, I. H. (2006). Is this happiness I see? Biases in the identification of emotional facial expressions in depression and social phobia. *J Abnorm Psychol*, *115*(4), 705-714. https://doi.org/10.1037/0021-843X.115.4.705

Leppänen, J., Milders, M., Bell, J., Terriere, E., & Hietanen, J. (2004). Depression biases the recognition of emotionally neutral faces. *Psychiatry Research*, *128*, 123-133. https://doi.org/10.1016/j.psychres.2004.05.020

Morey, R. D., & Rouder, J. N. (2023). BayesFactor: Computation of Bayes Factors for Common Designs. https://CRAN.R-project.org/package=BayesFactor

Münkler, P., Rothkirch, M., Dalati, Y., Schmack, K., & Sterzer, P. (2015). Biased Recognition of Facial Affect in Patients with Major Depressive Disorder Reflects Clinical State. *PLOS ONE*, *10*(6), e0129863. https://doi.org/10.1371/journal.pone.0129863

Paiva-Silva, A. I. d., Pontes, M. K., Aguiar, J. S. R., & de Souza, W. C. (2016). How do we evaluate facial emotion recognition? *Psychology & neuroscience*, *9*(2), 153.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825-2830.

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychol Bull*, *115*(1), 102-141. https://doi.org/10.1037/0033-2909.115.1.102

Schurgin, M. W., Nelson, J., Iida, S., Ohira, H., Chiao, J. Y., & Franconeri, S. L. (2014). Eye movements during emotion recognition in faces. *Journal of Vision*, *14*(13), 14-14. https://doi.org/10.1167/14.13.14

Silvia, P. J., Allan, W. D., Beauchamp, D. L., Maschauer, E. L., & Workman, J. O. (2006). Biased Recognition of Happy Facial Expressions in Social Anxiety. *Journal of Social and Clinical Psychology*, *25*(6), 585-602. https://doi.org/10.1521/jscp.2006.25.6.585

Surguladze, S. A., Young, A. W., Senior, C., Brebion, G., Travis, M. J., & Phillips, M. L. (2004). Recognition accuracy and response bias to happy and sad facial expressions in patients with major depression. *Neuropsychology*, *18*(2), 212-218. https://doi.org/10.1037/0894-4105.18.2.212

Sweeny, T. D., Grabowecky, M., Suzuki, S., & Paller, K. A. (2009). Long-lasting effects of subliminal affective priming from facial expressions. *Consciousness and Cognition*, *18*(4), 929-938. https://doi.org/https://doi.org/10.1016/j.concog.2009.07.011

Testerman, M. (2023). *Face Image Databases: An A-Z directory of databases of face stimulus images for use in behavioral research*. Retrieved 2024/02/07 from https://libguides.princeton.edu/facedatabases

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., Marcus, D. J., Westerlund, A., Casey, B. J., & Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, *168*(3), 242-249. https://doi.org/https://doi.org/10.1016/j.psychres.2008.05.006

Vikhanova, A., Mareschal, I., & Tibber, M. (2022). Emotion recognition bias depends on stimulus morphing strategy. *Attention, Perception, & Psychophysics*, *84*(6), 2051-2059. https://doi.org/10.3758/s13414-022-02532-0