Measuring emotion perception ability using AI-generated stimuli

Ben Weidmann and Yixian Xu

We present a new measure of emotion perception called PAGE (Perceiving AI Generated Emotions). The test includes 20 emotions, expressed by ethnically diverse faces, spanning a wide range of ages. We created stimuli with Generative AI, illustrating a method to build customizable assessments of emotional intelligence at relatively low cost. Study 1 describes the validation of the image set and test construction. Study 2 reports the psychometric properties of the test, including convergent validity and relatively strong reliability. Study 3 explores predictive validity using a lab experiment in which we causally identify the contributions managers make to teams. PAGE scores predict managers' causal contributions to group success, a finding which is robust to controlling for personality and demographic characteristics. We discuss the potential of Generative AI to automate development of non-cognitive skill assessments.

Author Affiliation: Harvard Kennedy School, Harvard University, Cambridge, MA 02138

Corresponding author: Ben Weidmann; benweidmann@hks.harvard.edu

Author Contributions: BW and YX designed research, performed research, analyzed data, and wrote the paper.

Competing Interest Statement: The authors declare no conflict of interest.

Classification: Social Sciences; Psychological and Cognitive Sciences

Keywords: Emotion perception; Teamwork; Management; Measurement; Generative AI

Significance Statement

Reading other people's emotions is a foundational skill for social interaction, teamwork and management. Tests that measure this skill are used in psychology, economics and medicine. But existing assessments suffer from limitations, such as only testing people's ability to perceive emotions on Caucasian faces. We develop a new test with diverse stimuli expressing a broad range of complex emotions. We believe that by providing access to a free, reliable, inclusive test, more researchers will be able to investigate the role emotional intelligence plays in a wide range of settings. We also hope that researchers will build on our approach of using generative AI to build better assessments of non-cognitive skills.

Main Text

1. Introduction

The ability to recognize the emotional states of others matters for social interaction and workplace success. Empirical research has shown that emotional perceptiveness predicts income, job performance (1, 2), leadership emergence (3), teamwork effectiveness (4–6) and successful negotiation outcomes (7, 8).

Researchers began developing skill-based tests to measure emotional perceptiveness in the 1970s. These tests generally ask participants to assess emotional expressions that are portrayed by actors in videos, images or audio recordings. The assessments have been used in a wide range of disciplines including psychology, economics and medicine. However, existing measures face four challenges that may limit their usefulness (summarized in Table 1). First, most tests use ethnically homogenous stimuli and often *only* include Caucasian faces (9–15). This can result in biased tests, as participants recognize emotions more quickly and accurately when the person expressing the emotions shares their cultural and ethnic identity (16, 17). Ethnically homogenous test stimuli are especially problematic when assessments are used in a diverse workforce. Second, many assessments lack emotional range and encompass only a handful of basic emotions – often including only one positive emotion (9, 11, 15, 18–20). This can lead to ceiling effects and also may limit the external validity of tests, as real-world contexts involve a wide range of complex emotions (21, 22). Third, many existing tests have practical limitations that make them difficult for researchers to use, including the length (9, 13–15, 23) and cost of tests (11, 18, 19), along with the lack of freedom to use assessments on whatever platform researchers find convenient (14). Finally, many existing emotion perception assessments have not yet demonstrated predictive validity in teamwork or workplace settings, limiting their empirical usefulness for management and teamwork research.

To address these limitations, our paper develops a practical test that includes a wide range of emotions expressed by racially diverse faces, spanning ages 20 to 60. We then examine the predictive validity of the test in the context of managing a team. Stimuli were created with Generative AI, providing some initial evidence that Generative AI tools may assist researchers in constructing customizable assessments of emotional intelligence at relatively low cost. Models such as OpenAI's DALL-E and Google's Imagen are capable of generating photorealistic images using simple text prompts (24, 25). We use this tool to develop our diverse emotion recognition assessment called 'Perceiving AI Generated Emotions (PAGE)'.

Research Overview. The rest of the paper proceeds as follows. Study 1 describes the construction of the test. Studies 2a and 2b assess its psychometric properties and examine convergent validity by reporting on the correlation between 'Perceiving AI Generated Emotions (PAGE)' and 'Reading the Mind in the Eyes Test (RMET)', a widely-used measure of emotion perception and theory of mind (10). Study 3 explores the predictive validity of the PAGE test. Using repeated random assignment of managers to groups, we examine the extent to which PAGE predicts the causal contribution that managers make to team success.

2. Study 1: Test Construction

Study 1 reports on the methods used to generate and validate the faces used in the PAGE instrument. We also describe other aspects of the test design including selection of distractors for the multiple-choice task and the scoring method.

2.1 Create Face Stimuli with Generative AI

Face stimuli were generated using DALL-E 2, a diffusion-based model that allows users to generate photorealistic images from text prompts (24, 26). We selected DALL-E 2, as diffusion-based models

have been shown to generate higher-quality facial images compared to GANs in previous studies (27).

Many emotion perception tasks are limited to the "basic 6 emotions" – anger, disgust, fear, happiness, sadness, and surprise (28). Using multiple emotion elicitation methods and prompt engineering techniques, we experimented with generating faces expressing 25 emotions, as recent research suggests that people can reliably recognize up to 28 emotions from facial-bodily expressions (29). We only include emotions that aren't dependent on bodily expression and contextual clues. The 25 emotions are six basic emotions (Happiness, Anger, Fear, Anger, Disgust, Surprise), and 19 complex emotions: Disappointment (30), Amusement (31), Anxiety (32), Awe (33), Boredom (12, 34, 35), Concentration, Confusion, Contemplation (36), Contempt (37), Contentment (30), Desire, Doubt (38), Embarrassment (31), Interest (39), Pain (40), Pride (41), Relief, Shame (42), Sympathy (43).

2.2 Validation and Selection of Stimuli

To validate the emotional expressions generated using DALL-E 2, we recruited 500 participants with diverse demographic backgrounds on Prolific[1] to rate the stimuli in emotion categories. We computed the proportion of participants selecting the target emotion of each stimuli (out of 25 possible options) and retained those where a clear plurality of people identified the target emotion. From this sample, we selected a subset of 35 images, with the goal of producing a test that included racially diverse stimuli, gender equity, and a wide range of complex and basic emotions (see *Appendix*, Table S1 for characteristics of the images).

2.3 Test Construction

We generated a multiple-choice question for each image by selecting five distractors. These distractors were drawn from two main sources: first, the emotion labels chosen by participants during the stimuli validation task; second, other plausibly relevant emotions from the 25 emotions included in the test (29). Additionally, emotions which are frequent parts of social interactions such as *confusion*, *doubt*, and *interest* (36, 38) are overrepresented in the distractors, which reflects a desire to aid the predictive validity of the PAGE in real-world settings – especially those requiring teamwork. A list of face stimuli, target emotions, and distractors for each item is available in *Appendix*, Table S2.

The resulting set of 35 test questions were sequenced such that consecutive items did not feature the same emotion. The placement of both the correct emotion and the distractors were randomized. Correct answers are scored as 1, incorrect as 0. All materials are freely available. We also made both the short and full versions of the PAGE task publicly accessible via our lab website. Detailed descriptions of the construction of the short PAGE, and both instruments are provided in the *Appendix*. See Figure 2 for an example item of the PAGE test.

3. Study 2a: Measurement Properties of PAGE

Results. We recruited a diverse group of 1010 participants on Prolific to complete the PAGE task. The mean score for PAGE is 23.7 (*SD* = 5.0). There was no evidence of ceiling or floor effects, and no statistically significant difference in task performance between men and women, as shown in Figure 3. There was a negative correlation between PAGE scores and age ($r = -0.14$, $p < 0.001$), as shown in Figure 4. These findings are consistent with previous work showing lower accuracy at emotional recognition in older adults (45, 46).

Item difficulty. The mean item difficulty – measured in terms of the proportion of people submitting correct answers – was 0.68 (SD = 0.12). Individual item difficulties ranged from 0.44 to 0.89, this range falls within recommended parameters for psychological assessment measures (47). Table 2

[1] Douglas et al. (44) compare several measures of data quality such as sample representativeness and find that Prolific is comparable or better than competitors such as MTurk.

presents the distribution of item difficulties. A full table of item difficulty across items and populations is available in Table S3 in *Appendix.*

Internal consistency. Cronbach's alpha for the PAGE test was 0.73, compared to an average reliability of α = 0.60 reported for other emotion recognition ability tests (Boone & Schlegel, 2016). This is notable given the brevity of the test – 8 minutes on average – and the wide range of emotions included in the test, both of which tend to reduce internal consistency.

Factor structure. Exploratory factor analysis suggests that the PAGE has a one-factor structure (*Appendix*, Figure S1). To further assess the unidimensionality of the PAGE instrument, we tested a one-factor confirmatory factor analysis (CFA) model. Model fit was evaluated by inspecting the comparative fit index (CFI), the Tucker–Lewis Index (TLI), and the root mean square error of approximation (RMSEA). The one-factor CFA model showed good fit (χ^2 = 884, df = 560, p < 0.001, CFI = 0.829, TLI = 0.818, RMSEA = 0.024). Although the CFI and TLI are slightly lower than common acceptability threshold (0.9), the low RMSEA and satisfactory values of Cronbach's alpha overall suggest that the PAGE is a unidimensional test. The factor structure of PAGE is consistent with findings about the dimensionality of emotion recognition abilities, that ERA is one broad ability consisting of correlated valence-based skills and minor ability facets related to pairs of similar and highly confused emotions (48, 49).

4. Study 2b: Convergent Validity of PAGE

Results. We examined the convergent validity of PAGE by estimating the correlation of PAGE scores with Reading the Mind in the Eyes Test (RMET), a widely used measure of emotion perception. We find that the PAGE is highly correlated with RMET (raw correlation = 0.66, disattenuated correlation = 0.88, p < 0.001), providing evidence of convergent validity. We also explored the patterns of performance between PAGE and RMET by age, gender, ethnicity, shown in Figure 4. The PAGE test shows the same performance patterns as RMET. First, there is no statistically meaningful difference in task performance between women and men. Second, both tests show stable performance from age 18-40, then a slight decline up to age 60, where our sample ends. Last, mean performance of Caucasian participants is slightly better on both tests.

5. Study 3: Predictive Validity of the PAGE assessment

To assess the predictive validity of PAGE we fielded the test in a lab experiment that used a novel design to identify the causal contribution that individual managers make to group performance. The experimental design – summarized in Figure 5 – randomly assigns managers to four different teams of workers. Our design makes use of the repeated random assignment of managers to teams to identify the average impact each manager has on group performance (5). A total of 116 managers in the experiment completed the PAGE instrument, which allowed us to compare PAGE scores with causally identified manager contributions. We find that PAGE scores are predictive of manager contributions, and that this association is robust to controls for personality and demographic factors. These participants also completed the Reading the Mind in the Eyes Test (RMET), providing a benchmark for comparison.

Results. We find that the PAGE score of each manager is positively associated with group performance. The correlation between manager PAGE scores and group performance = 0.189 (p < 0.001, df = 406). These results are consistent with findings from non-hierarchical teamwork settings where team members' emotion perception ability positively predicts team success (5, 7).

Next, we move from the group level to the level of individual managers. We identify and estimate the average *causal* impact each manager has by exploiting the fact managers are randomly assigned to multiple teams (5). We find that PAGE scores positively predict the average causal contribution that managers have on their groups (correlation = 0.290, p =0.002, df = 113).

Table 3 contextualizes PAGE's association with managerial causal contributions, by comparing its predictive validity to RMET. Overall, we find that the association between PAGE scores and

manager contributions is greater than that for RMET, and that the association is robust to controls for differences in Big 5 personality measures and demographic factors (age, gender, ethnicity and education). Column 1 (Table 3) presents the raw association between manager causal contribution and PAGE scores, which are standardized to have mean = 0 and SD = 1. We find that a 1sd increase in PAGE scores is associated with an increase in manager contributions of 0.290sd. Columns 2 and 3 add controls for Big 5 personality and demographics which reduces the coefficients, but they remain significant: after controlling for age, ethnicity, gender, education program and Big 5 measures, a 1sd change in PAGE is associated with a 0.230sd increase in managerial contributions (p=0.031). Columns 4 to 6 repeat this process, focusing on RMET as a predictor. The relationship between RMET and manager causal contributions are weaker, and not statistically significant. Column 7 is a full specification in which we include all variables, illustrating the robust association between PAGE scores and the impact that managers have on their teams.

Why might PAGE be more predictive of managerial performance than RMET? One possibility is that the emotions in the PAGE assessment were chosen from a list of emotions that can be clearly expressed by faces (29). In contrast, some of the target emotions in RMET may be better characterized as dispositions (e.g. 'cautious') that are difficult to express and perceive in faces. We also explored the possibility that the predictive performance of PAGE was enhanced by the fact that Study 3's participant sample is ethnically diverse and may have benefited from PAGE's more diverse set of stimuli. To examine this empirically we split teams into two sets: ethnically homogenous (all group members self-identify with the same ethnicity) and ethnically diverse (at least two group members self-identify with different ethnicities). For each set we separately calculate the association between team performance and the manager's score on RMET|PAGE. Among diverse groups, the PAGE test significantly predicted performance ($\hat{\rho}$ = 0.216, p < 0.001, n=342). This was not the case for homogenous groups ($\hat{\rho} = -0.002$, p=0.985, n=66). The correlation between RMET and group performance was not statistically significant for either homogeneous or diverse groups.

Finally, we return to the question of how a manager's emotional perceptiveness might improve team performance. As noted above, one of the roles of the manager is to motivate their teammates. Motivation matters in Study 3's collaborative task for two reasons. First, participants are given a large number of cognitively demanding puzzles that require effort to solve. Second, 'workers' in the experiment do not receive financial incentives based on performance, so they are prone to lose interest over the course of the task. As noted in Weidmann et al. (2024), the task involves three periods of intensive problem solving, divided by two dedicated breaks in which managers can take stock and motivate their team.

To explore the role that motivation plays in the task, we separately calculate the causal contribution that managers make in each of the three problem-solving periods (i.e. start; middle; end). Weidmann et al., (2024) find that the last period matters most in terms of managerial contribution. We extend this finding by documenting that the PAGE test is most strongly predictive of performance in the final period of the task, as noted in Table 4. Column 1 in the table regresses each manager's average causal contribution *during the first period of the task* against their PAGE score. Column 2 repeats the exercise, focusing on the average causal contribution managers make during the middle period of the task. Column 3 presents results for the final period. A one standard-deviation increase in the PAGE test is associated with a 0.235 standard-deviation increase in managerial contributions in the final period (n=87, p=0.022). The association between PAGE and managerial contributions in the first and second periods are not statistically significant.

6. Discussion and Conclusion

This paper develops and validates a measure of emotional perceptiveness using a demographically diverse set of 35 faces, expressing 20 emotions. The PAGE test materials are open source. Study 1 shows that generative AI is capable of producing standardized, realistic faces that express both basic and complex emotions. Studies 2a and 2b demonstrate the psychometric properties of the test including unidimensionality, internal validity and convergent validity. Finally, Study 3 provides

evidence of predictive validity, especially for researchers interested in leadership and management, by showing that PAGE scores are associated with the causal contribution that managers make to group performance in a controlled lab study. Overall, these results suggest that the PAGE may be useful for researchers looking for a short, skill-based test of emotional perceptiveness that is suited for studies of teamwork and management among demographically diverse populations.

The PAGE assessment also illustrates the possibility that generative AI can help create customized measures of emotional intelligence by substantially reducing the cost of test creation and automating the test development process. There are at least two ways in which tests may be usefully customized. First, researchers may find it helpful to be able to vary the demographic profile of the stimuli. For example, researchers working with a sample of elderly adults may want a test in which stimuli have older faces than are found in existing assessments. Second, it may be beneficial to have tests that oversample specific complex emotions – many of which are absent from most measures. For example, an organization hiring a team leader may screen for the ability to recognize confusion*,* as this potentially enables quick clarification. In contexts where teamwork is important, perceiving anxiety may signal a colleague's need for support. Of course, it is an empirical question whether such customized measures of emotional recognition are more predictive of positive outcomes in real-world contexts, but with the advent of generative AI this research agenda is much more practically achievable.

Improvements in AI technology may further reduce the costs of creating tests of emotional perceptiveness. We manually created multiple-choice questions for the PAGE test. However, generative AI is now capable of creating multiple-choice questions across difficulty levels (50). More significantly, recent research suggests the possibility that large language models (LLMs) may be able to produce similar results to human participants in social science research (51). If this is true of emotional perceptivity, test developers would be able to combine human responses with a low-cost sample of LLM respondents to assess and refine the psychometric properties of new tests. Overall, while existing AI technology reduced the practical barriers we faced in creating PAGE, it seems likely that these barriers will be progressively lowered.

In closing, we believe that the PAGE test measures a general construct that is an important determinant of success in a wide range of social activities, from negotiation and hiring, to networking and working in a team. The test has strong measurement properties, is appropriate for diverse populations, and is open access. We hope that others will build on the approach of using generative AI to create and validate customized tests that allow for a better understanding of the role emotion plays in facilitating interaction in the workplace and beyond.

7. Materials and Methods

7.1 Study 1: Test Construction

7.1.1 Prompt Engineering to Generate Emotional Faces

The prompts we gave DALL-E 2 used three methods, derived from emotion elicitation strategies researchers use when creating emotional stimuli using human actors. The first method is simply to instruct expressers to express a particular emotion (as used, for example, in 'Karolinska Directed Emotional Faces (KDEF)' and NimStim (52, 53)), e.g. *'a 22 year old Caucasian woman feeling very angry.*' The second method relies on the Directed Facial Action Task (54), in which expressers are instructed to employ specific facial actions based on the emotion prototypes identified by Ekman and colleagues (55). For example, to express the emotion pride we prompted DALL-E 2 by saying *'a 30 year old Asian man showing pride. His head is held high, jaw thrust out, he has a small smile, lip pressed.'* Finally, we borrow a technique from studies of cultural variation in emotional expressions and use a short story to elicit emotions in expressers (56). For example, to generate an image of surprise, the prompt includes the following text *'a 47 year old Indian woman showing a surprised face when hearing a loud sound she didn't expect.*'

We experimented with a combination of these three methods to elicit emotions in AI images, operationalized as prompts in three formats: 'emotion word', 'emotion word, and facial actions'

'emotion word, and one-sentence emotion story.' To generate human-like images, each prompt begins with 'Generate a photorealistic image of…'. We also added 'detailed skin texture', and 'proportional eyes' into the prompt, which are found to be among the key factors in making an AI face look more realistic (57). An example prompt that uses all three of the techniques listed above is as follows: *'A realistic photo of a 20 year old Indian woman caught embarrassed and blushing in a social gaffe. Her whole face and head are in the middle. Plain grey background (leave some blank space around). She is wearing a white t-shirt. No body language, head oriented at the front, and staring at the camera.'* A full list of prompts is provided for each emotion in Appendix, Table S4.

7.1.2 Stimuli Standardization

We generated 150 realistic faces for the initial stim set. These faces represent 25 emotions, six ethnicities, and ages ranging from 20 to 60. We used the same ethnicity categories as the Chicago Face Database: Asian, Black, Caucasian, Indian, Latino, Multi-racial (58). We then used Adobe Photoshop to standardize the stimuli to have consistent grey background. Images were resized so that each target's face and head are in the middle. See Figure 1 for sample stimuli.

7.1.3 Participant recruitment and validation of the stimulus set

We recruited 500 participants on Prolific to rate the stimuli in emotion categories. Participants provided written informed consent to take part in the study with data collected solely for research purposes. Each participant rated between 30 and 35 images. Participants were asked to select one emotion that best described the face, from a list of 25 emotions. Each image was rated by at least 100 participants. Our sample was ethnically diverse and displayed gender balanced (Female 49%, Mean age 34 years, White 57.8%, see *Appendix*, Table S5 for full demographic statistics). Each participant received a compensation of $2 for completion of the task.

7.2 Study 2a: Measurement Properties of PAGE

Participants and procedure. We recruited 1010 participants from Prolific. Participants provided written informed consent to take part in the study with data collected solely for research purposes. All participants were located in the United States and were ethnically diverse (Female 50%, Mean age 36.7 years, White 44.5%, see *Appendix*, Table S5 for demographic details). We oversampled non-Caucasian participants so that we could better assess performance on the PAGE of people from different ethnic backgrounds. We also focused on respondents who were full-time workers (91%) aged 25-55, to validate the PAGE among a sample that would be of use to labor economists and organizational psychologists. We administered the PAGE test on Qualtrics (Full instructions in *Appendix*) and each participant received a compensation of $2.50. To motivate participants to maintain attention throughout the task, we also awarded the quintile of performers a $5 bonus. The median participant spent 8 minutes on the test.

7.3 Study 2b: Convergent Validity of PAGE

Reading the Mind in the Eyes test (RMET). To demonstrate convergent validity, we compare results on the PAGE test to the Reading the Mind in the Eyes Test (RMET). The RMET is a 36-item multiple choice test measuring emotional perceptiveness by presenting cropped images of faces that only include the eye region (Baron-Cohen et al., 2001).

Participants and procedure. We analyze a sub-sample of 741 participants from Study 2a, who completed both PAGE and RMET, administered on Qualtrics. Participant demographics are presented in *Appendix*, Table S5. Both tasks included one practice question to familiarize participants with the task format. To limit the effect of differences in vocabulary, we provided a list of emotion definitions for reference. When participants put their cursor above the emotion word, they were provided with a definition. To reduce the impact of order effects, we had 249 participants complete the RMET first (then the PAGE) and 492 people complete the tests in the reverse order.

Question order was also randomized for both tests. On average, participants completed the PAGE in 8 minutes and the RMET in 10 minutes.

7.4 Study 3: Predictive Validity of PAGE

Participants. We recruited an ethnically diverse sample of graduate and undergraduate students at the University of Essex in the UK. Participants provided written informed consent to take part in the study. The median participant was 25 years old and had 2 years' work experience (*Appendix*, Table S5). Participants were paid £29 for completing the study, with a performance bonus that ranged from £0 to £12. The average payment was £35.

Experiment procedures. The goal of the experiment was to causally identify the contribution that managers made to teams, and to explore the characteristics that were predictive of management performance. The experiment included both individual and group assessments (see **Error! Reference source not found.**4 for experiment overview). Individual tests included a demographic questionnaire and a shortened version of the Big 5 inventory (59). A subsample of experimental participants (n=116 managers) completed the PAGE assessment as well as the Reading the Mind in the Eyes Test (RMET).

Group testing took place at Essex Lab in England. Each team consists of a manager and two workers. Teams work face-to-face on a collaborative task that aims to emulate some of the core demands of real-world hierarchical teams by requiring managers to co-ordinate, monitor and motivate workers. The task is described in detail in (Weidmann et al., 2024). Briefly, teams are asked to simultaneously make progress on three different 'modules'. A central responsibility of the managers is to make decisions about which module each member of the team – including themselves – should work on. Decisions about task allocations are fully dynamic and managers can change them at any time. The group task takes around 15 minutes in total and includes dedicated time for managers to introduce themselves and motivate their team. Talking is allowed throughout the task. After each group finished the task, managers are randomly assigned to another team. Over the course of the experiment each manager is randomly assigned to four groups.

To succeed in the task managers have to assign their teammates tasks that match their skills, monitor their performance and maintain high levels of effort and engagement (5). We hypothesize that managers who are more skilled at perceiving emotions are better placed to meet these demands. For example, a manager skilled at detecting boredom may be faster at perceiving low morale and better able to respond by providing encouragement or switching the task a teammate is working on. Similarly, a manager who can perceive that a teammate is confused is in a better position to offer support and avoid their teammate submitting incorrect solutions on behalf of the team.

Data, Materials, and Software Availability. The data and code for all studies, AI-generated images, and experiment materials will be publicly available through an Open Science Foundation repository (DOI 10.17605/OSF.IO/7A4XS).

Acknowledgments. We would like to gratefully acknowledge the financial support of Schmidt Futures.

References

1. S. Côté, Emotional Intelligence in Organizations. *Annual Review of Organizational Psychology and Organizational Behavior* **1**, 459–488 (2014).

2. H. A. Elfenbein, Emotion in Organizations: Theory and Research. *Annu. Rev. Psychol.* **74**, 489–517 (2023).

3. S. Côté, P. N. Lopes, P. Salovey, C. T. H. Miners, Emotional intelligence and leadership emergence in small groups. *The Leadership Quarterly* **21**, 496–508 (2010).

4. C. I. C. C. Farh, M.-G. Seo, P. E. Tesluk, Emotional intelligence, teamwork effectiveness, and job performance: The moderating role of job context: Journal of Applied Psychology. *Journal of Applied Psychology* **97**, 890–900 (2012).

5. B. Weidmann, D. J. Deming, Team Players: How Social Skills Improve Team Performance. *Econometrica* **89**, 2637–2657 (2021).

6. C. Riedl, Y. J. Kim, P. Gupta, T. W. Malone, A. W. Woolley, Quantifying collective intelligence in human groups. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2005737118 (2021).

7. H. A. Elfenbein, M. D. Foo, J. White, H. H. Tan, V. C. Aik, Reading your Counterpart: The Benefit of Emotion Recognition Accuracy for Effectiveness in Negotiation. *J Nonverbal Behav* **31**, 205–223 (2007).

8. K. Schlegel, M. Mehu, J. M. van Peer, K. R. Scherer, Sense and sensibility: The role of cognitive and emotional intelligence in negotiation. *Journal of Research in Personality* **74**, 6–15 (2018).

9. M. Bell, G. Bryson, P. Lysaker, Positive and negative affect recognition in schizophrenia: a comparison with substance abuse and normal control subjects. *Psychiatry Research* **73**, 73–82 (1997).

10. S. Baron-Cohen, S. Wheelwright, J. Hill, Y. Raste, I. Plumb, The “Reading the Mind in the Eyes” Test Revised Version: A Study with Normal Adults, and Adults with Asperger Syndrome or High-functioning Autism. *Child Psychology Psychiatry* **42**, 241–251 (2001).

11. J. D. Mayer, P. Salovey, D. R. Caruso, G. Sitarenios, Measuring emotional intelligence with the MSCEIT V2.0. *Emotion* **3**, 97–105 (2003).

12. K. R. Scherer, H. Ellgring, Multimodal expression of emotion: Affect programs or componential appraisal patterns? *Emotion* **7**, 158–171 (2007).

13. T. Bänziger, K. R. Scherer, J. A. Hall, R. Rosenthal, Introducing the MiniPONS: A Short Multichannel Version of the Profile of Nonverbal Sensitivity (PONS). *J Nonverbal Behav* **35**, 189–204 (2011).

14. K. Schlegel, K. R. Scherer, Introducing a short version of the Geneva Emotion Recognition Test (GERT-S): Psychometric properties and construct validation. *Behav Res* **48**, 1383–1392 (2016).

15. K. R. Scherer, U. Scherer, Assessing the Ability to Recognize Facial and Vocal Expressions of Emotion: Construction and Validation of the Emotion Recognition Index. *J Nonverbal Behav* **35**, 305–326 (2011).

16. H. A. Elfenbein, N. Ambady, Universals and Cultural Differences in Recognizing Emotions. *Curr Dir Psychol Sci* **12**, 159–164 (2003).

17. S. G. Young, K. Hugenberg, M. J. Bernstein, D. F. Sacco, Perception and motivation in face recognition: a critical review of theories of the Cross-Race Effect. *Pers Soc Psychol Rev* **16**, 116–142 (2012).

18. S. Nowicki, M. P. Duke, Individual differences in the nonverbal communication of affect: The diagnostic analysis of nonverbal accuracy scale. *J Nonverbal Behav* **18**, 9–35 (1994).

19. D. Matsumoto, *et al.*, A New Test to Measure Emotion Recognition Ability: Matsumoto and Ekman's Japanese and Caucasian Brief Affect Recognition Test (JACBART). *Journal of Nonverbal Behavior* **24**, 179–209 (2000).

20. C. G. Kohler, *et al.*, Facial Emotion Recognition in Schizophrenia: Intensity Effects and Error Pattern. *AJP* **160**, 1768–1774 (2003).

21. G. A. van Kleef, *The Interpersonal Dynamics of Emotion* (Cambridge University Press, 2016).

22. L. H. Phillips, G. Slessor, Moving Beyond Basic Emotions in Aging Research. *J Nonverbal Behav* **35**, 279–286 (2011).

23. T. Bänziger, D. Grandjean, K. R. Scherer, Emotion recognition from expressions in face, voice, and body: The Multimodal Emotion Recognition Test (MERT). *Emotion* **9**, 691–704 (2009).

24. A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, M. Chen, Hierarchical Text-Conditional Image Generation with CLIP Latents. [Preprint] (2022). Available at: http://arxiv.org/abs/2204.06125 [Accessed 24 April 2024].

25. C. Saharia, *et al.*, Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. [Preprint] (2022). Available at: http://arxiv.org/abs/2205.11487 [Accessed 24 April 2024].

26. G. Marcus, E. Davis, S. Aaronson, A very preliminary analysis of DALL-E 2. (2022). https://doi.org/10.48550/ARXIV.2204.13807.

27. M. Stypułkowski, *et al.*, Diffused Heads: Diffusion Models Beat GANs on Talking-Face Generation. [Preprint] (2023). Available at: http://arxiv.org/abs/2301.03396 [Accessed 24 April 2024].

28. P. Ekman, E. R. Sorenson, W. V. Friesen, Pan-Cultural Elements in Facial Displays of Emotion. *Science* **164**, 86–88 (1969).

29. A. S. Cowen, D. Keltner, What the face displays: Mapping 28 emotions conveyed by naturalistic expression. *American Psychologist* **75**, 349–364 (2020).

30. D. T. Cordaro, M. Brackett, L. Glass, C. L. Anderson, Contentment: Perceived Completeness across Cultures and Traditions. *Review of General Psychology* **20**, 221–235 (2016).

31. D. Keltner, Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame. *Journal of Personality and Social Psychology* **68**, 441–454 (1995).

32. A. M. Perkins, S. L. Inchley-Mort, A. D. Pickering, P. J. Corr, A. P. Burgess, A facial expression for anxiety. *Journal of Personality and Social Psychology* **102**, 910–924 (2012).

33. M. N. Shiota, *et al.*, Beyond happiness: Building a science of discrete positive emotions. *American Psychologist* **72**, 617–643 (2017).

34. K. R. Scherer, H. Ellgring, Are facial expressions of emotion produced by categorical affect programs or dynamically driven by appraisal? *Emotion* **7**, 113–130 (2007).

35. D. T. Cordaro, *et al.*, The recognition of 18 facial-bodily expressions across nine cultures. *Emotion* **20**, 1292–1300 (2020).

36. P. Rozin, A. B. Cohen, High frequency of facial expressions corresponding to confusion, concentration, and worry in an analysis of naturally occurring facial expressions of Americans. *Emotion* **3**, 68–75 (2003).

37. D. Matsumoto, P. Ekman, The Relationship Among Expressions, Labels, and Descriptions of Contempt. *Journal of Personality and Social Psychology* **87**, 529–540 (2004).

38. C. F. Benitez-Quiroz, R. B. Wilbur, A. M. Martinez, The not face: A grammaticalization of facial expressions of emotion. *Cognition* **150**, 77–84 (2016).

39. J. Reeve, The face of interest. *Motiv Emot* **17**, 353–375 (1993).

40. K. M. Prkachin, The consistency of facial expressions of pain: a comparison across modalities. *Pain* **51**, 297–306 (1992).

41. J. L. Tracy, R. W. Robins, Show Your Pride: Evidence for a Discrete Emotion Expression. *Psychol Sci* **15**, 194–197 (2004).

42. J. L. Tracy, D. Matsumoto, The spontaneous expression of pride and shame: Evidence for biologically innate nonverbal displays. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 11655–11660 (2008).

43. J. L. Goetz, D. Keltner, E. Simon-Thomas, Compassion: An evolutionary analysis and empirical review. *Psychological Bulletin* **136**, 351–374 (2010).

44. B. D. Douglas, P. J. Ewell, M. Brauer, Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. *PLOS ONE* **18**, e0279720 (2023).

45. A. Mill, J. Allik, A. Realo, R. Valk, Age-related differences in emotion recognition ability: A cross-sectional study. *Emotion* **9**, 619–630 (2009).

46. T. Ruffman, J. D. Henry, V. Livingstone, L. H. Phillips, A meta-analytic review of emotion recognition and aging: Implications for neuropsychological models of aging. *Neuroscience & Biobehavioral Reviews* **32**, 863–881 (2008).

47. R. M. Thorndike, T. M. Thorndike-Christ, *Measurement and Evaluation in Psychology and Education, 8th Edition* (Pearson, 2010).

48. J. A. Hall, “The PONS test and the psychometric approach to measuring interpersonal sensitivity” in *Interpersonal Sensitivity: Theory and Measurement*, (Lawrence Erlbaum Associates Publishers, 2001), pp. 143–160.

49. K. Schlegel, D. Grandjean, K. R. Scherer, Emotion recognition: Unidimensional ability or a set of modality- and emotion-specific skills? *Personality and Individual Differences* **53**, 16–21 (2012).

50. V. Aryadoust, A. Zakaria, Y. Jia, Investigating the affordances of OpenAI's large language model in developing listening assessments. *Computers and Education: Artificial Intelligence* **6**, 100204 (2024).

51. J. J. Horton, Large Language Models as Simulated Economic Agents: What Can We Learn from Homo Silicus? [Preprint] (2023). Available at: http://arxiv.org/abs/2301.07543 [Accessed 10 May 2024].

52. D. Lundqvist, A. Flykt, A. Öhman, Karolinska Directed Emotional Faces. https://doi.org/10.1037/t27732-000. Deposited 1998.

53. N. Tottenham, *et al.*, The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research* **168**, 242–249 (2009).

54. P. Ekman, "The Directed Facial Action Task:Emotional Responses Without Appraisal" in *Handbook of Emotion Elicitation and Assessment*, J. A. Coan, J. J. B. Allen, Eds. (Oxford University PressNew York, NY, 2007), pp. 47–53.

55. P. Ekman, Are there basic emotions? *Psychological Review* **99**, 550–553 (1992).

56. D. T. Cordaro, *et al.*, Universals and cultural variations in 22 emotional expressions across five cultures: Emotion. *Emotion* **18**, 75–93 (2018).

57. E. J. Miller, *et al.*, AI Hyperrealism: Why AI Faces Are Perceived as More Real Than Human Ones. *Psychol Sci* **34**, 1390–1403 (2023).

58. D. S. Ma, J. Correll, B. Wittenbrink, The Chicago face database: A free stimulus set of faces and norming data. *Behav Res* **47**, 1122–1135 (2015).

59. S. D. Gosling, P. J. Rentfrow, W. B. Swann, A very brief measure of the Big-Five personality domains. *Journal of Research in Personality* **37**, 504–528 (2003).

60. B. Weidmann, J. Vecci, F. Said, D. J. Deming, S. R. Bhalotra, How Do You Find a Good Manager? [Preprint] (2024). Available at: https://www.nber.org/papers/w32699 [Accessed 22 July 2024].

61. H. A. Kim, *et al.*, Multiracial Reading the Mind in the Eyes Test (MRMET): An inclusive version of an influential measure. *Behavior Research Methods* **56**, 5900 (2024).

Figures and Tables

Fig. 1 Study 1: Example Stimuli from the PAGE test, generated using DALL-E 2. In this case, the emotions being represented by the four stimuli are: a) surprise, b) contentment, c) anger, d) fear

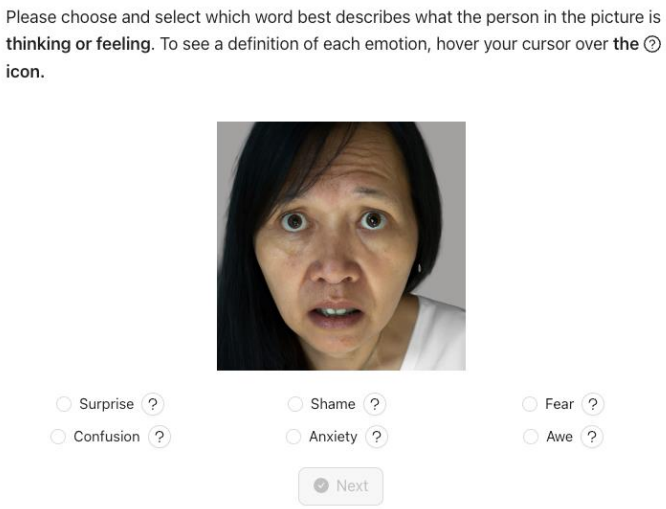


Fig. 1 Study 1: Sample item from the PAGE test. For each item, participants select one answer from six options. Definitions of emotions are provided to participants if they click on the question marks. The target emotion for this item is *fear*.

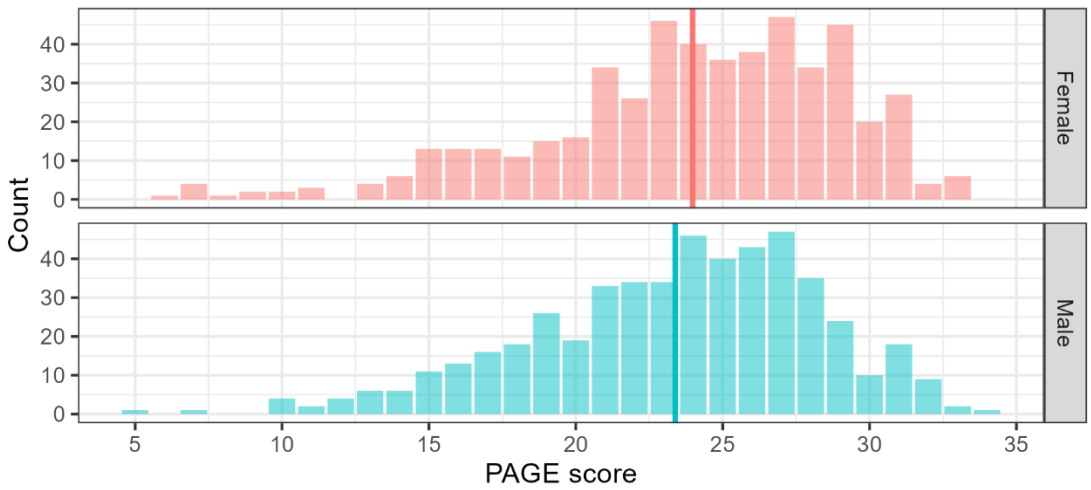


Fig. 3 Study 2a: Distribution of PAGE scores by gender (N = 505 males and 505 females). Mean score for male is 23.4 (*SD* = 4.9), Mean score for female is 24.0 (*SD* = 5.1).

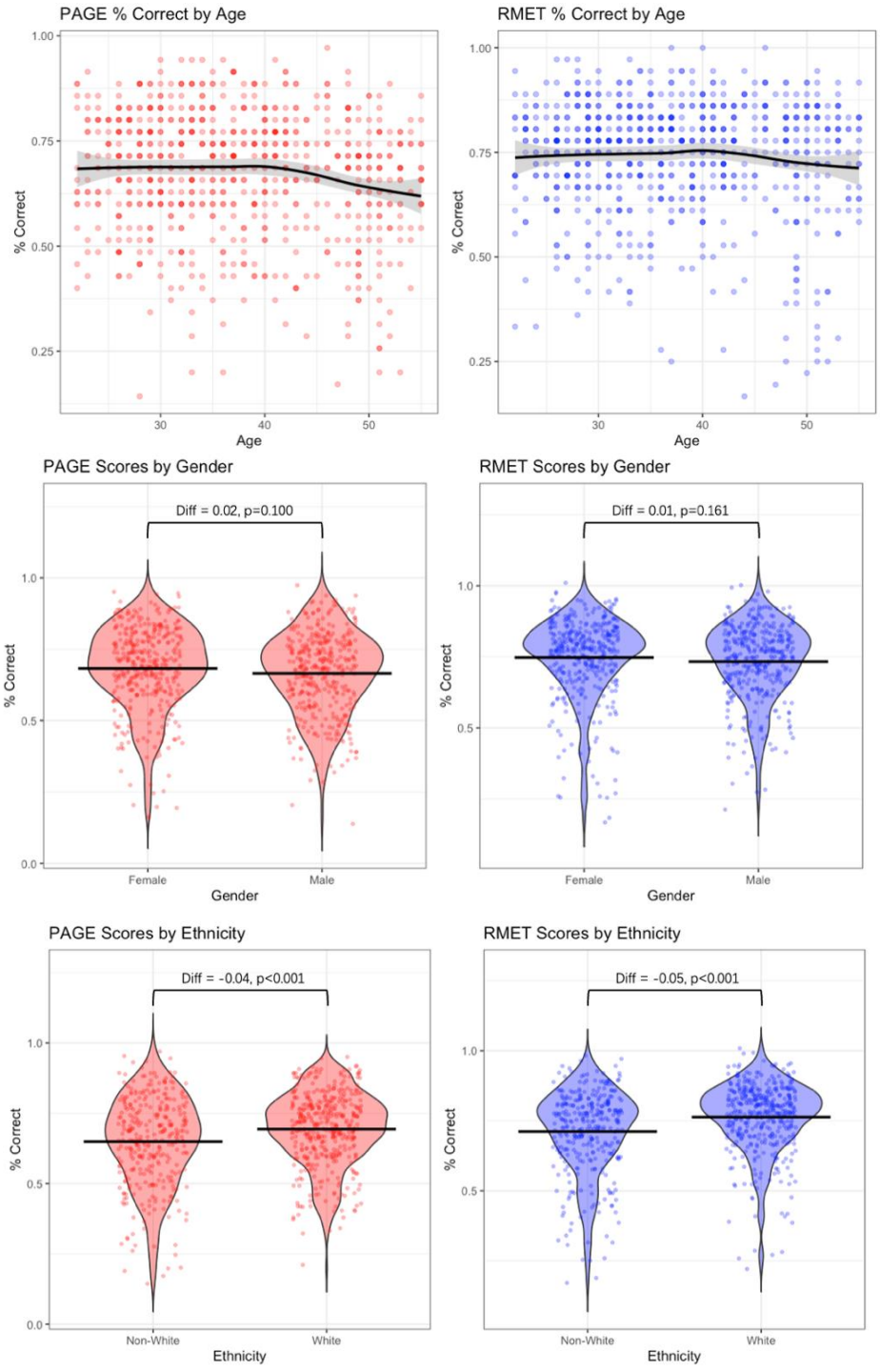


Fig. 4 Study 2b: Patterns of performance for PAGE/RMET by gender/age/ethnicity (N = 741). 50% of the participants are female, and 50.5% are White. Both tests show similar performance patterns across gender, age and ethnicity.

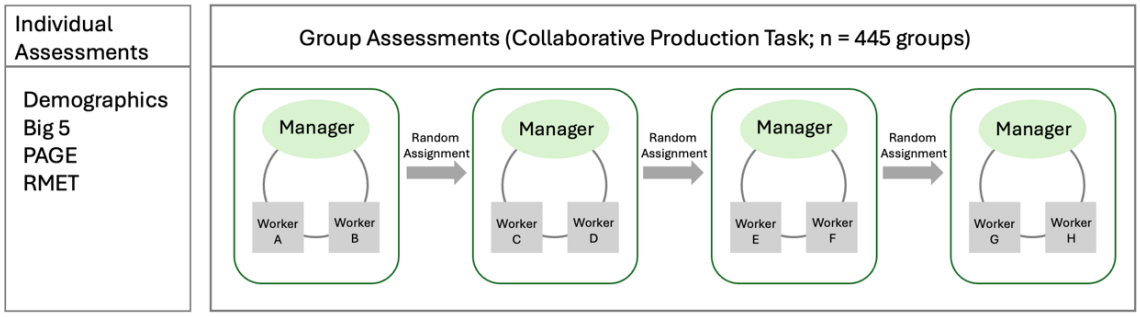


Fig. 5 Study 3 experiment overview. Participants were randomly assigned to the role of 'manager' or 'worker'. Each group completed a novel collaborative problem-solving task in which managers assigned tasks, monitored group progress, and motivated teammates. After each task, managers were randomly reassigned to new groups, managing a total of four different groups. The repeated random assignment of managers to teams enabled the identification of each manager's contribution to group performance.

Table 1. Existing measures of emotional perception and their potential limitations.

Test	Emotional Range	Ethnic Diversity	Practical Challenges	Item #
DANVA-2 (18)	4 emotions	Caucasian, Black	Not freely available	48
BLERT (9)	7 emotions	Caucasian	15 – 20 minutes	21
JACBART (19)	7 emotions	Asian, Caucasian	Not freely available	56
RMET (10)	26 mental states	Caucasian	None	36
PERT-96 (20)	5 emotions	Diverse	None	96
MSCEIT Perception Tests (11)	5 emotions	Caucasian	Not freely available	50
MERT (23)	10 emotions	Caucasian	45 - 60 minutes	120
MiniPONS (13)	2 affective situations	Caucasian	15 – 20 minutes	64
ERI (15)	5 emotions	Caucasian	15 - 20 minutes	60
GERT-S (14)	14 emotions	Caucasian	15 - 20 minutes; No customization	42
MRMET (61)	18 mental states	Diverse	None	37 or 10

Table notes. DANVA: Diagnostic Assessment of Non-Verbal Abilities; BLERT: Bell Lysaker Emotion Recognition Task; JACBART: Japanese and Caucasian Brief Affect Recognition Test; RMET: The Reading the Mind in the Eyes Test; PERT-96: Penn Emotion Recognition Task; MSCEIT: Mayer–Salovey–Caruso Emotional Intelligence Test; MERT: Multimodal Emotion Recognition Test; MiniPONS: Profile of Nonverbal Sensitivity (short version); ERI: Emotion Recognition Index; GERT-S: Geneva Emotion Recognition Test (short version); MRMET: Multiracial Reading the Mind in the Eyes Test.

Table 2. Distribution of Item Difficulty

Difficulty Range	Number of Items
$0.30 \leq p < 0.50$	3 (8.6%)
$0.50 \leq p < 0.70$	18 (51.4%)
$0.70 \leq p < 0.90$	14 (40%)

Table notes. 'Difficulty range' indicates the range of proportion of correct responses ('p').

Table 3. Associations between manager performance and emotional perceptiveness measured by PAGE vs. RMET

	Average Causal Contributions of Managers						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
PAGE	0.290**	0.244*	0.230*				0.273*
	(0.090)	(0.094)	(0.105)				(0.113)
RMET				0.149	0.074	-0.025	-0.125
				(0.089)	(0.092)	(0.111)	(0.116)
Big5		X	X		X	X	X
Demographics			X			X	X
Observations	115	111	109	115	111	109	109
R^2	0.084	0.172	0.223	0.024	0.124	0.184	0.233
Adjusted R^2	0.076	0.125	0.117	0.016	0.073	0.073	0.118

Table notes. *p<0.05, **p<0.01, ***p<0.001. The dependent variable is "manager's estimated causal contribution" as measured by the average score across each manager's randomly assigned teams. Demographic factors include age, ethnicity, education and gender. PAGE and RMET scores are both standardized to have mean = 0 and SD = 1.

Table 4. does emotional perceptiveness matter more at the start or end of the task?

	First period (1)	Second period (2)	Final period (3)
PAGE	0.120	0.085	0.235*
	(0.116)	(0.102)	(0.101)
Constant	-0.003	-0.013	0.043
	(0.113)	(0.100)	(0.099)
Observations	87	87	87
R^2	0.013	0.008	0.060

Table notes. *p<0.05, **p<0.01, ***p<0.001. The dependent variable for each regression is the causal contribution of managers for a specific period of the game (n=88 managers). Column (1) represents the first third of the game; column (2) the middle period; and column (3) the last period. The dependent variable and the PAGE variable are standardized to have mean=0 and sd=1.

Supporting Information for

Measure emotion perception ability using AI-generated stimuli

Authors: Ben Weidmann, Yixian Xu

Author Affiliation: Harvard Kennedy School, Harvard University, Cambridge, MA 02138

Corresponding author: Ben Weidmann

Corresponding author email: benweidmann@hks.harvard.edu

This PDF file includes:

Figures S1

Tables S1 to S5

SI References

Other supporting materials for this manuscript include the following:

PAGE task (Study 2a, Study 2b)

RMET task (Study 2b)

Big 5 inventory (Study 3)

Demographics survey (Study 3)

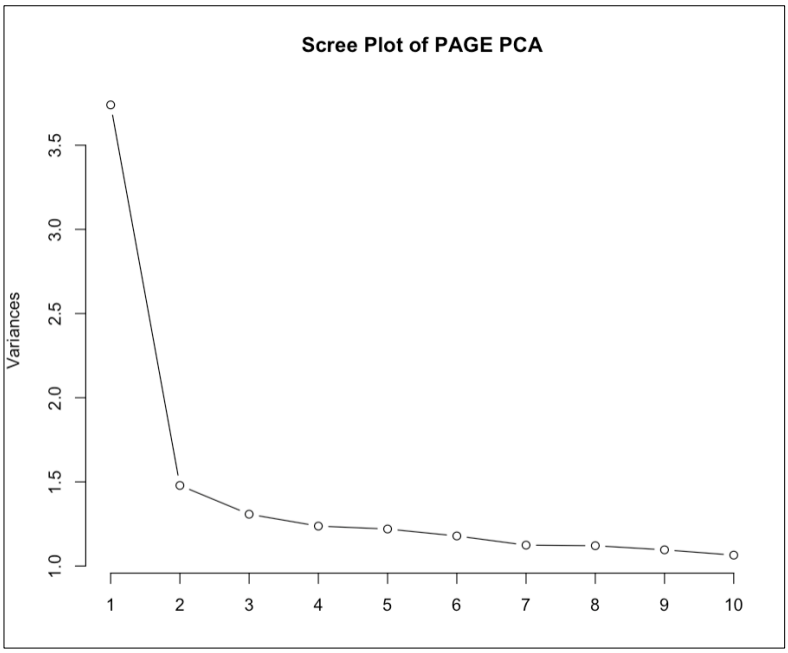
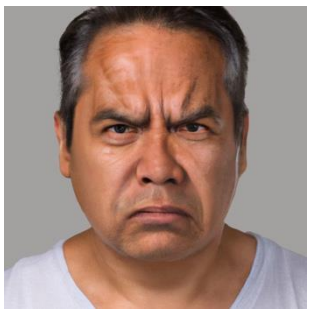


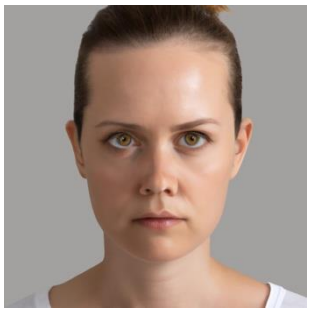
Fig. S1 Study 2a: Scree plot of PAGE. The scree plot displays the eigenvalues associated with each factor extracted during exploratory factor analysis (EFA). The steep drop between the first and second factors suggests that the PAGE test has a dominant one-factor structure.

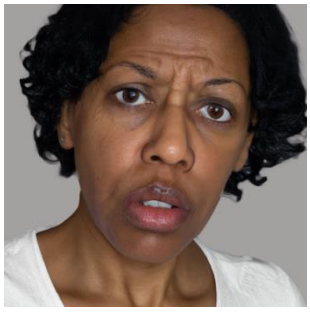
Table S1 PAGE image count by demographics

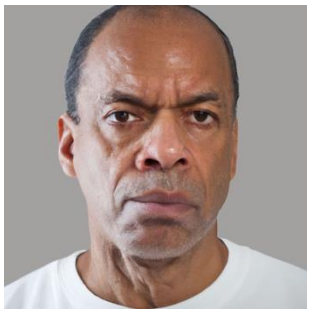
Ethnicity	#	Age	#	Gender	#
Caucasian	11	20-29	5	Female	17
Black	8	30-39	16	Male	18
Latino	9	40-59	13	Total	35
Asian	4	60	1		
Indian	2	Total	35		
Multi-racial	1				
Total	35				

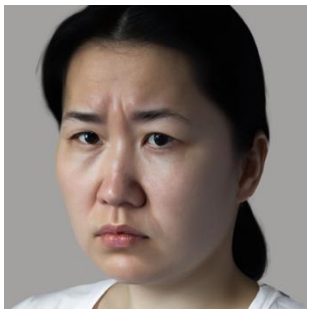
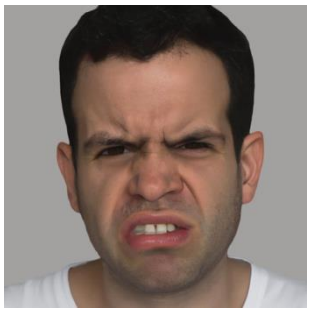
Table S2 PAGE stimuli, target emotions, and distractors

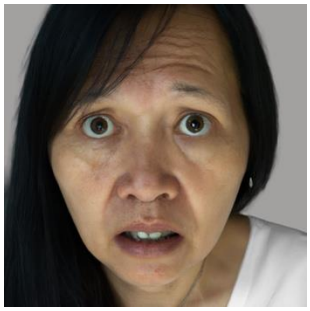
#	Stimuli	Target emotion	Distractors in multiple-choice test
1		Amusement	Awe, Pleasure, Interest, Surprise, Relief
2		Anger	Pride, Pain, Disgust, Confusion, Shame
3		Anger	Pride, Pain, Confusion, Sadness, Disgust
4		Anger	Amusement, Contempt, Sadness, Disappointment, Doubt

5		Anxiety	Contentment, Embarrassment, Contemplation, Confusion, Fear
6		Anxiety	Contentment, Disappointment, Disgust, Relief, Boredom
7		Boredom	Interest, Distress, Pleasure, Pain, Anger
8		Concentration	Confusion, Doubt, Contentment, Disappointment, Interest

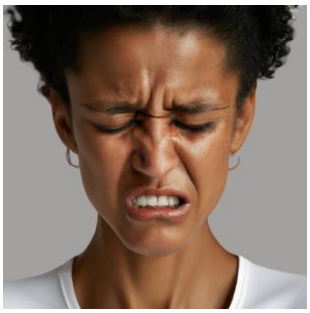
9		Concentration	Contentment, Interest, Contempt, Anger, Disappointment
10		Confusion	Surprise, Interest, Anxiety, Doubt, Disgust
11		Contemplation	Confusion, Surprise, Disappointment, Interest, Contentment
12		Contemplation	Anxiety, Relief, Surprise, Interest, Contentment

13		Contempt	Anxiety, Disgust, Confusion, Interest, Boredom
14		Contentment	Ecstasy, Pride, Desire, Contemplation, Contempt
15		Contentment	Ecstasy, Disappointment, Pride, Contempt, Relief
16		Contentment	Amusement, Pride, Desire, Joy, Contempt

17		Disappointment	Confusion, Anger, Disgust, Contempt, Boredom
18		Disgust	Amusement, Pain, Contempt, Confusion, Anger
19		Disgust	Contemplation, Pain, Contempt, Confusion, Anger
20		Doubt	Contentment, Confusion, Anger, Contemplation, Interest

21		Doubt	Interest, Confusion, Boredom, Sadness, Anxiety
22		Embarrassment	Relief, Confusion, Pride, Anxiety, Shame
23		Fear	Awe, Confusion, Surprise, Anxiety, Shame
24		Fear	Anger, Confusion, Awe, Embarrassment, Pleasure

25		Interest	Amusement, Boredom, Doubt, Contemplation, Sympathy
26		Interest	Doubt, Boredom, Embarrassment, Surprise, Disappointment
27		Joy	Awe, Surprise, Desire, Contentment, Confusion
28		Joy	Surprise, Contentment, Desire, Confusion, Awe

29		Pain	Pride, Disappointment, Anger, Embarrassment, Shame
30		Pride	Amusement, Awe, Interest, Joy, Contempt
31		Pride	Amusement, Doubt, Interest, Contentment, Contempt
32		Sadness	Contentment, Anxiety, Confusion, Pain, Boredom

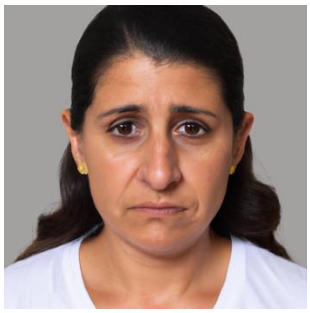
33		Sadness	Contentment, Anxiety, Disappointment, Pain, Boredom
34		Surprise	Interest, Ecstasy, Confusion, Anger, Fear
35		Surprise	Interest, Anger, Relief, Confusion, Disgust

Table S3 Item Difficulty Table

Item	Overall (n=1010)	Female	Male	White	Non_ White
Sadness_30_Caucasian_Male	87.2	87.4	87.1	88.9	85.9
Joy_45_Latino_Female_1	79.2	74.4	84.1	81.1	77.7
Confusion_40_Black_Female	67.2	73.4	61	71.7	63.6
Anger_30_Black_Male	84.2	83.6	84.7	86	82.7
Surprise_47_Indian_Female	89.1	89	89.3	90	88.4
Disgust_35_Caucasian_Male	74.1	76.7	71.4	76.6	72
Fear_50_Asian_Female	70	71.2	68.8	73.5	67.2
Boredom_25_ Caucasian _Female	64.9	59.2	70.6	61	67.9
Anxiety_40_Black_Male	64.6	63.1	66	66.8	62.7
Doubt_30_Latino_Female_1	65.7	65.1	66.4	63.3	67.7
Disappointment_30_Asian_Female	61.4	62.7	60	62.6	60.4
Interest_50_ Latino _Male	61.8	63.7	59.8	63.5	60.4
Embarrassment_20_Indian_Female	44.6	48.7	40.4	48.8	41.2
Pride_30_Asian_Male	66.6	65.7	67.6	73.3	61.3
Contentment_55_Black_Male	54.4	61.5	47.1	58.1	51.3
Fear_50_Caucasian_Male	71.1	71.4	70.8	75.7	67.4
Pride_30_Caucasian_Female	66.1	70.2	62	72.4	61.1
Contentment_35_ Latino _Male	64.8	70	59.4	69.9	60.6
Amusement_35_Caucasian_Male	61.8	63.7	59.8	64.1	59.9
Anxiety_34_Caucasian_Male	44.2	42.6	45.7	43.2	44.9
Contempt_50_Black_Male	47.7	49.9	45.5	51.2	44.9
Anger_22_Caucasian_Female	84.8	86	83.5	83.7	85.6
Contentment_45_Caucasian_Female	64.4	70.2	58.4	67.9	61.5
Sadness_30_Latino_Female	74.4	73.4	75.3	78.4	71.1
Doubt_30_Latino_Female _2	65.9	66.5	65.4	68.2	64.2
Surprise_30_Asian_Female	88.6	86.4	90.9	90.6	87
Interest_25_Latino _Male	59.3	59.8	58.8	61.7	57.4
Pain_30_Black_Female	61.9	63.5	60.2	62.4	61.5
Concentration_25_Caucasian_Female	70.7	70.6	70.8	71.5	70.1
Anger_40_Latino _Male	81.2	81.5	80.9	82.2	80.4
Joy_45_Latino _Female_2	72.7	68.6	76.7	73.7	71.8
Contemplation_30_Black_Male_1	50.8	53.6	47.9	50.1	51.3
Concentration_60_Multi-racial_Male	64.3	64.1	64.4	61.2	66.7
Disgust_45_Caucasian_Male	70.5	69.6	71.4	73.1	68.4
Contemplation_30_Black_Male_2	68.5	70.8	66.2	69.5	67.7

Note. The item labels are written in the format of 'emotion_age_ethnicity_gender' to indicate the emotion and demographic information of each face. For example, 'Sadness_30_Caucasian_Male' represents a face of 30 year old male expressing the emotion sadness.

Table S4 Emotion Prompts for PAGE Stimulus Generation

Emotion (Sources)	Prompt	Method
Amusement (1)	"Generate a photorealistic image of a [age] [ethnicity] [gender] laughing with jaw dropping, head tilting backwards, with detailed skin texture and natural lighting, with highly realistic, well-proportioned eyes, with opened eyes. No body language, showing the face and shoulder, head oriented at the front, and looking at the camera. Plain grey background, wearing a white t-shirt."	facial actions
Anger (2)	"A realistic photo of a [age] [ethnicity] [gender] feeling very **angry**. Symmetric eyes. No body language, face in the middle, head oriented at the front, and staring at the camera. Plain grey background, wearing a white t-shirt"	emotion word
Anxiety (3)	"Create a hyper-realistic image of a [age] [ethnicity] [gender] showing expression **anxiety**. Eyes looking sideways, frowned eyebrows, biting lips. Detailed skin texture and natural lighting. Wearing a white t-shirt. No body language, showing only the face, head oriented at the front, and staring at the camera. Plain grey background."	emotion word + facial actions
Boredom (4–6)	"Create a hyper-realistic image of a [age] [ethnicity] [gender] showing expression **boredom**. Eyelids dropping. Detailed skin texture and natural lighting. Wearing a white t-shirt. No body language, showing only the face, head oriented at the front, and staring at the camera. Plain grey background."	emotion word + facial actions
Concentration (7)	" A realistic photo of a [age] [ethnicity] [gender] feeling very **concentrated**, clearly paying attention to something intently. No body language, face in the middle, head oriented at the front, and staring at the camera. Plain grey background, wearing a white t-shirt No body language, face in the middle, head oriented at the front, . Plain grey background, wearing a white t-shirt"	emotion word + emotion story
Confusion (7)	"Create a hyper-realistic image of a [age] [ethnicity] [gender] showing a **confused** expression with slightly opened mouth. Detailed skin texture and natural lighting. No body language, showing only the face, head oriented at the front, and staring at the camera. Plain grey background, he is wearing a white t-shirt."	emotion word + facial actions
Contemplation (7)	"A realistic photo of a [age] [ethnicity] [gender] expressing the emotion **contemplation**, he is pondering life. No body language, face in the middle, head oriented at the front, and staring at the camera. Plain grey background, wearing a white t-shirt"	emotion word + emotion story

Contempt (8)	“Create a hyper-realistic photo of a [age] [ethnicity] [gender] expressing **contempt**. Detailed skin texture and natural lighting. No body language, showing only the face, head oriented at the front, and staring at the camera. Plain grey background, wearing a white T-shirt.”	emotion word
Contentment (9)	“A realistic photo of a [age] [ethnicity] [gender] experiencing a feeling of **well-being and delight**. His whole face and head in the middle. Plain grey background (leave some blank space around). He is wearing a white t-shirt. No body language, head oriented at the front, and staring at the camera.”	Synonym of emotion word
Disappointment (9)	“Create a hyper-realistic image of a [age] [ethnicity] [gender] showing expression **disappointment**. Eyebrows slightly furrowed, lips pressed, eyes looking sideways. Detailed skin texture and natural lighting. Wearing a white t-shirt. No body language, showing only the face, head oriented at the front, and staring at the camera. Plain grey background.”	emotion word + facial actions
Disgust (2)	“A realistic photo of a [age] [ethnicity] [gender] feeling **disgusted**. No body language, face in the middle, head oriented at the front, and staring at the camera. Plain grey background, wearing a white t-shirt”	emotion word
Doubt (10)	“Create a hyper-realistic image of a [age] [ethnicity] [gender] showing a **doubtful** expression with pressed lips. Detailed skin texture and natural lighting. No body language, showing only the face, head oriented at the front, and staring at the camera. Plain grey background, he is wearing a white t-shirt.”	emotion word + facial actions
Embarrassment (1)	“A realistic photo of a [age] [ethnicity] [gender] caught **embarrassed** and blushing in a social gaffe. Her whole face and head in the middle. Plain grey background (leave some blank space around). She is wearing a white t-shirt. No body language, head oriented at the front, and staring at the camera.”	emotion word + facial action + emotion story
Fear (2)	“Create a hyper-realistic image of a [age] [ethnicity] [gender] showing expression **fear**. Wearing a white t-shirt. Detailed skin texture and natural lighting. No body language, showing only the face, head oriented at the front, and staring at the camera. Plain grey background,	emotion word
Interest (11)	“Create a hyper-realistic image of a [age] [ethnicity] [gender] showing expression **interest**. His eyebrows pulled straight up, eyes open wide, he has a small smile, his head tilts forward,. Detailed skin texture and natural lighting. No body language, showing only the face, head oriented at the front, and staring at the camera. Plain grey background, he is wearing a white t-shirt.”	emotion word + facial action

Joy (2)	“A realistic photo of a [age] [ethnicity] [gender] expressing emotion **joy**, she is very happy at something unexpected. No body language, showing only the face, head oriented at the front, and staring at the camera. Plain grey background, wearing a white t-shirt””	emotion word + emotion story
Pain (12)	“Create a hyper-realistic image of a [age] [ethnicity] [gender] showing a **painful** expression. Her eyes closed tightly, her lips tighten and pressed. Detailed skin texture and natural lighting. No body language, showing only the face, head oriented at the front, and staring at the camera. Plain grey background, he is wearing a white t-shirt.””	emotion word + facial action
Pride (13)	“Create a hyper-realistic image of a [age] [ethnicity] [gender] showing **pride**. His head holds high, jaw thrusts out, he has a small smile, lip pressed. Detailed skin texture and natural lighting. No body language, showing only the face, head oriented at the front, and staring at the camera. Plain grey background, wearing a white t-shirt.”	emotion word + facial action
Sadness (2)	“A realistic photo of a [age] [ethnicity] [gender] showing a **sad** face when hearing an old friend's death. her whole face and head in the middle. Plain grey background (leave some blank space around). she is wearing a white t-shirt. No body language, face in the middle, head oriented at the front, and staring at the camera.”	emotion word + emotion story
Surprise (2)	“A realistic photo of a [age] [ethnicity] [gender] showing a **surprised** face when hearing something she didn't expect. her whole face and head in the middle. Plain grey background (leave some blank space around). she is wearing a white t-shirt. No body language, face in the middle, head oriented at the front, and staring at the camera.”	emotion word + emotion story

Table S5 Demographics of participants

	Study 1	Study 2a	Study 2b^	Study 3*
Number	500	1010	741	116
Ethnicity (%)				
White	57.8	44.5	50.5	17.0
Black/African American	5.8	22.3	21.5	16.1
Latino/Hispanicº	3.4	15.4	13.2	-
Asian°	20.0	17.9	14.8	56.2
Other / not reported	0.0	0.0	0.0	10.7
Age				
Mean (SD)	34.0 (9.4)	36.7 (9.1)	37.6 (9.1)	25.4 (4.5)
18-29 (%)	41.8	26.4	23.6	83.6
30-39 (%)	32.6	36.6	35.2	16.4
40-59 (%)	25.2	36.9	41.2	-
60-74(%)	0.4	-	-	-
Female (%)	49	50	50	43
Full-time workers (%)	46	91	100	-
Country	US	US	US	UK

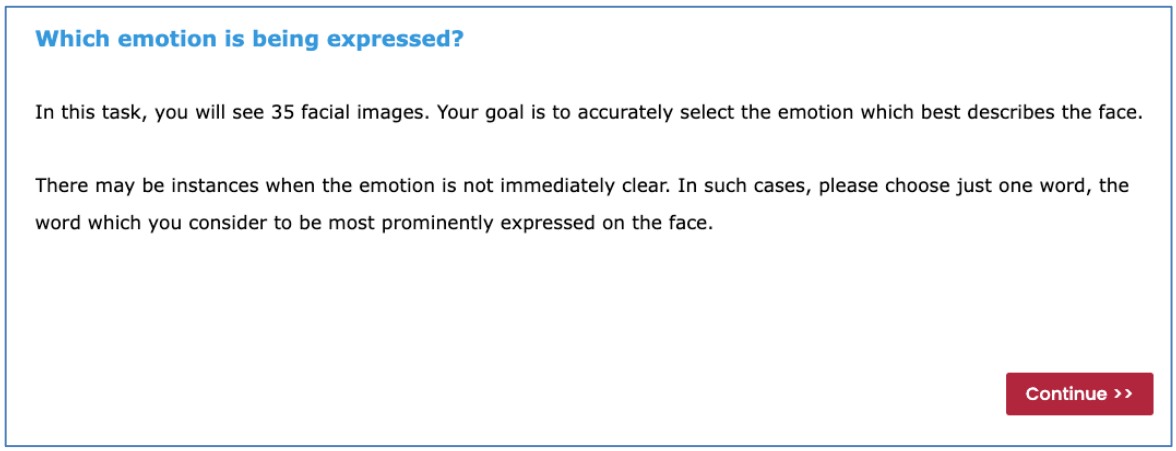
Note. ^Study 2b is a subset of study 2a. *This is the sample of managers in Study 3. The full experiment contained 555 participants. °This includes 'Asian British'; ºIn Study 3, which was done in the UK, this was not an option.

PAGE (short and long versions), RMET, and surveys

Construction of the short PAGE

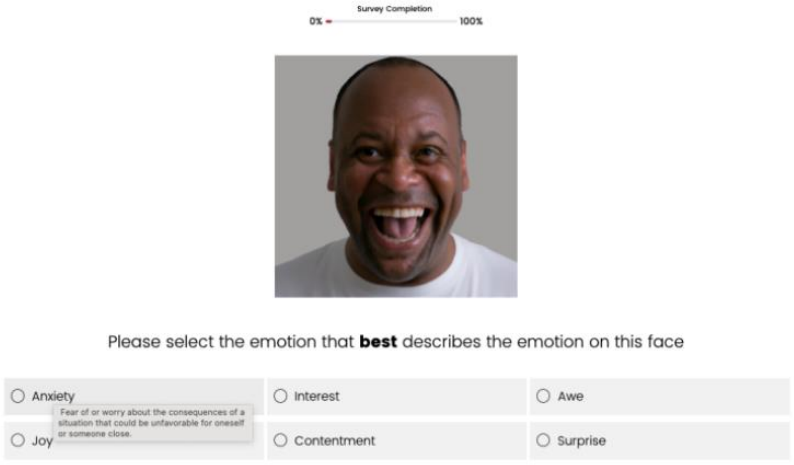
To construct the short version of PAGE, we ranked items by their correlation with the total test score, selecting higher-ranking items while ensuring facial diversity. We balanced gender (8 male, 8 female), age (20-60 years), and ethnicity (six categories), and included both basic and complex emotions. The resulting 16-item version PAGE takes only 4 minutes to complete, offering a more practical option for time-constrained experiments. Items included in the short version are: 2, 5, 8, 9, 10, 12, 16, 18, 20, 23, 24, 26, 27, 29, 30, 34.

PAGE task instructions



Under each image, 6 emotion words are presented. Before making your choice, make sure that you have read all 6 words. To know the definition of each emotion word, hover your cursor over **the specific word** for seconds and its definition will appear. See the example screenshot.

You can also look it up in the list of definitions. Please open this link in a new tab to reference during the survey. Note that not all emotions listed are expressed on these faces.



Most people take 6-8 minutes to complete the test.

Now you need to answer one simple practice questions to get familiar with the task format. Please take as much time as you need.

Continue >>

Please select the emotion that **best** describes the emotion on this face

◯ Surprise	◯ Contentment	◯ Awe
◯ Interest	◯ Anxiety	◯ Joy

Continue >>

PAGE items 1 – 35

Please select the emotion that **best** describes the emotion on this face

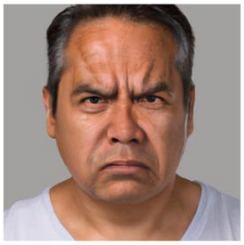
○ Confusion	○ Anger	○ Fear
○ Surprise	○ Ecstasy	○ Interest

Continue >>

Please select the emotion that **best** describes the emotion on this face

○ Contentment	○ Pain	○ Sadness
○ Boredom	○ Confusion	○ Anxiety

Continue >>



Please select the emotion that **best** describes the emotion on this face

○ Contempt	○ Sadness	○ Disappointment
○ Anger	○ Amusement	○ Doubt

Continue >>

Please select the emotion that **best** describes the emotion on this face

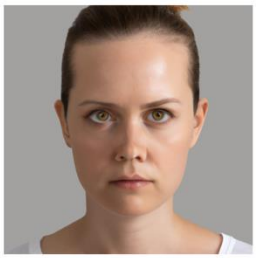
○ Confusion	○ Embarrassment	○ Pleasure
○ Anger	○ Fear	○ Awe

Continue >>

Please select the emotion that **best** describes the emotion on this face

○ Confusion	○ Awe	○ Desire
○ Joy	○ Contentment	○ Surprise

Continue >>



Please select the emotion that **best** describes the emotion on this face

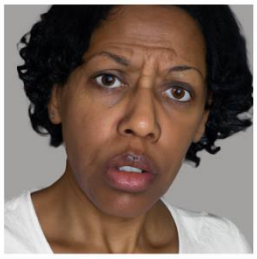
○ Confusion	○ Interest	○ Disappointment
○ Concentration	○ Contentment	○ Doubt

Continue >>

Please select the emotion that **best** describes the emotion on this face

○ Disgust	○ Contempt	○ Anger
○ Contemplation	○ Pain	○ Confusion

Continue >>



Please select the emotion that **best** describes the emotion on this face

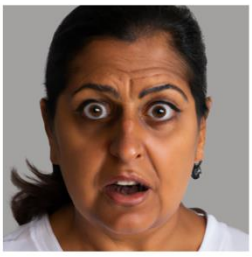
○ Surprise	○ Anxiety	○ Interest
○ Doubt	○ Disgust	○ Confusion

Continue >>

Please select the emotion that **best** describes the emotion on this face

○ Confusion	○ Anger	○ Sadness
○ Disgust	○ Pain	○ Pride

Continue >>



Please select the emotion that **best** describes the emotion on this face

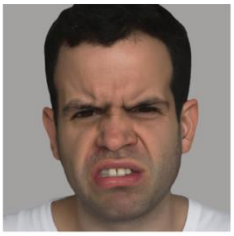
○ Relief	○ Disgust	○ Confusion
○ Anger	○ Surprise	○ Interest

Continue >>

Please select the emotion that **best** describes the emotion on this face

○ Anxiety	○ Contentment	○ Interest
○ Surprise	○ Contemplation	○ Relief

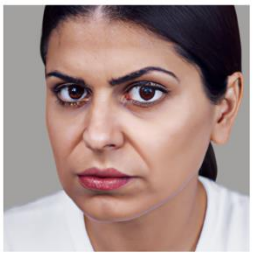
Continue >>



Please select the emotion that **best** describes the emotion on this face

○ Confusion	○ Pain	○ Anger
○ Disgust	○ Contempt	○ Amusement

Continue >>



Please select the emotion that **best** describes the emotion on this face

○ Sadness	○ Interest	○ Anxiety
○ Boredom	○ Confusion	○ Doubt

Continue >>

Please select the emotion that **best** describes the emotion on this face

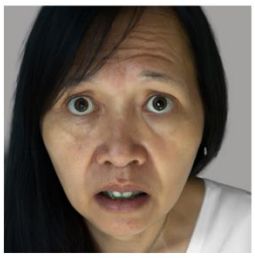
○ Contemplation	○ Desire	○ Ecstasy
○ Contempt	○ Contentment	○ Pride

Continue >>

Please select the emotion that **best** describes the emotion on this face

○ Contempt	○ Doubt	○ Pride
○ Amusement	○ Interest	○ Contentment

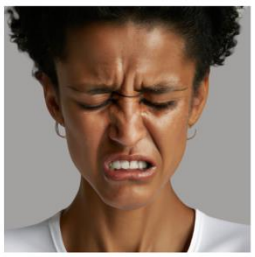
Continue >>



Please select the emotion that **best** describes the emotion on this face

○ Shame	○ Awe	○ Confusion
○ Surprise	○ Fear	○ Anxiety

Continue >>



Please select the emotion that **best** describes the emotion on this face

○ Pain	○ Anger	○ Shame
○ Embarrassment	○ Pride	○ Disappointment

Continue >>

Please select the emotion that **best** describes the emotion on this face

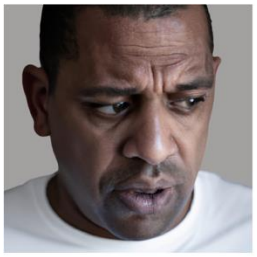
○ Anger	○ Boredom	○ Interest
○ Distress	○ Pain	○ Pleasure

Continue >>

Please select the emotion that **best** describes the emotion on this face

○ Surprise	○ Awe	○ Interest
○ Amusement	○ Pleasure	○ Relief

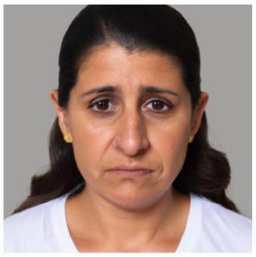
Continue >>



Please select the emotion that **best** describes the emotion on this face

○ Anxiety	○ Boredom	○ Disgust
○ Relief	○ Contentment	○ Disappointment

Continue >>



Please select the emotion that **best** describes the emotion on this face

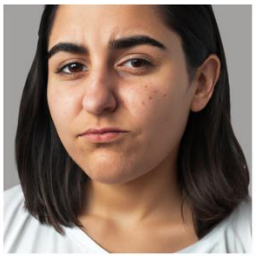
○ Disappointment	○ Sadness	○ Pain
○ Boredom	○ Contentment	○ Anxiety

Continue >>

Please select the emotion that **best** describes the emotion on this face

○ Contentment	○ Ecstasy	○ Pride
○ Disappointment	○ Relief	○ Contempt

Continue >>



Please select the emotion that **best** describes the emotion on this face

○ Interest	○ Anger	○ Contemplation
○ Confusion	○ Doubt	○ Contentment

Continue >>

Please select the emotion that **best** describes the emotion on this face

○ Concentration	○ Contempt	○ Disappointment
○ Anger	○ Contentment	○ Interest

Continue >>

Please select the emotion that **best** describes the emotion on this face

○ Shame	○ Confusion	○ Anger
○ Disgust	○ Pride	○ Pain

Continue >>

Please select the emotion that **best** describes the emotion on this face

◯ Boredom	◯ Contempt	◯ Disgust
◯ Anger	◯ Confusion	◯ Disappointment

Continue >>

Please select the emotion that **best** describes the emotion on this face

◯ Boredom	◯ Surprise	◯ Disappointment
◯ Interest	◯ Doubt	◯ Embarrassment

Continue >>

Please select the emotion that **best** describes the emotion on this face

○ Interest	○ Confusion	○ Contentment
○ Contemplation	○ Surprise	○ Disappointment

Continue >>

Please select the emotion that **best** describes the emotion on this face

○ Anxiety	○ Pride	○ Confusion
○ Embarrassment	○ Shame	○ Relief

Continue >>

Please select the emotion that **best** describes the emotion on this face

○ Interest	○ Pride	○ Awe
○ Joy	○ Contempt	○ Amusement

Continue >>

Please select the emotion that **best** describes the emotion on this face

○ Confusion	○ Joy	○ Contentment
○ Awe	○ Surprise	○ Desire

Continue >>

Please select the emotion that **best** describes the emotion on this face

◯ Anxiety	◯ Contentment	◯ Contemplation
◯ Embarrassment	◯ Fear	◯ Confusion

Continue >>

Please select the emotion that **best** describes the emotion on this face

◯ Boredom	◯ Disgust	◯ Confusion
◯ Contempt	◯ Interest	◯ Anxiety

Continue >>

Please select the emotion that **best** describes the emotion on this face

◯ Contempt	◯ Contentment	◯ Amusement
◯ Desire	◯ Pride	◯ Joy

Continue >>

Please select the emotion that **best** describes the emotion on this face

◯ Boredom	◯ Doubt	◯ Amusement
◯ Sympathy	◯ Contemplation	◯ Interest

Continue >>

RMET Task instructions

Please re-enter your Prolific ID

Continue >>

Task 1: Read the Mind in the Eyes

On the following pages you will be presented with a series of 36 pictures of eyes. For each set of eyes, choose and select which word best describes what the person in the picture is **thinking or feeling**.

You may feel that more than one word is applicable but please choose just one word, the word which you consider to be most suitable.

Continue >>

Task 1: Read the Mind in the Eyes

On the following pages you will be presented with a series of 36 pictures of eyes. For each set of eyes, choose and select which word best describes what the person in the picture is **thinking or feeling**.

You may feel that more than one word is applicable but please choose just one word, the word which you consider to be most suitable.

Continue >>

Before making your choice, make sure that you have read all 4 words. To know the meaning of each emotion word, hover your cursor over **the specific word** for seconds and its definition, and one example sentence will appear. See the example screenshot.

You can also look it up in the list of definitions. Please open this link in a new tab to reference during the survey. Note that not all words listed are expressed by these eyes. Please try your best, thanks!

Please choose and select which word best describes what the person in the picture is **thinking or feeling.**

jealous	panicked	arrogant	hateful
envious. Tony was jealous of all the taller, better-looking boys in his class.	○	○	○

Continue >>

Practice

Please choose and select which word best describes what the person in the picture is **thinking or feeling.**

jealous	panicked	arrogant	hateful
○	○	○	○

Continue >>

RMET items 1 – 36

Please choose and select which word best describes what the person in the picture is **thinking or feeling.**

doubtful	aghast	affectionate	playful
○	○	○	○

Continue >>

Please choose and select which word best describes what the person in the picture is **thinking or feeling.**

irritated	disappointed	depressed	accusing
○	○	○	○

Continue >>

Please choose and select which word best describes what the person in the picture is **thinking or feeling.**

contemplative	flustered	encouraging	amused
○	○	○	○

Continue >>

Please choose and select which word best describes what the person in the picture is **thinking or feeling.**

preoccupied	annoyed	hostile	horrified
○	○	○	○

Continue >>

Please choose and select which word best describes what the person in the picture is **thinking or feeling.**

pensive	hostile	irritated	excited
○	○	○	○

Continue >>

Please choose and select which word best describes what the person in the picture is **thinking or feeling.**

apologetic	defiant	contented	curious
○	○	○	○

Continue >>

Please choose and select which word best describes what the person in the picture is **thinking or feeling.**

terrified	baffled	distrustful	aghast
○	○	○	○

Continue >>

Please choose and select which word best describes what the person in the picture is **thinking or feeling.**

embarrassed	skeptical	indifferent	dispirited
○	○	○	○

Continue >>

Please choose and select which word best describes what the person in the picture is **thinking or feeling.**

sympathetic	thoughtful	encouraging	irritated
○	○	○	○

Continue >>

Please choose and select which word best describes what the person in the picture is **thinking or feeling.**

aghast	bored	cautious	insisting
○	○	○	○

Continue >>

Please choose and select which word best describes what the person in the picture is **thinking or feeling.**

alarmed	bewildered	serious	ashamed
○	○	○	○

Continue >>

Please choose and select which word best describes what the person in the picture is **thinking or feeling.**

aghast	decisive	amused	bored
○	○	○	○

Continue >>

Please choose and select which word best describes what the person in the picture is **thinking or feeling.**

despondent	relieved	excited	shy
○	○	○	○

Continue >>

Please choose and select which word best describes what the person in the picture is **thinking or feeling.**

dominant	friendly	guilty	horrified
○	○	○	○

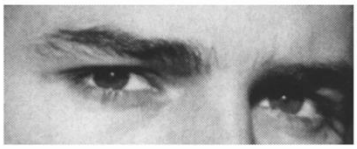
Continue >>

Please choose and select which word best describes what the person in the picture is **thinking or feeling.**

grateful	arrogant	sarcastic	tentative
○	○	○	○

Continue >>

Please choose and select which word best describes what the person in the picture is **thinking or feeling.**



embarrassed	fantasizing	concerned	guilty
○	○	○	○

Continue >>

Please choose and select which word best describes what the person in the picture is **thinking or feeling.**

aghast	reflective	irritated	impatient
○	○	○	○

Continue >>

Please choose and select which word best describes what the person in the picture is **thinking or feeling.**

grateful	insisting	imploring	preoccupied
○	○	○	○

Continue >>

Please choose and select which word best describes what the person in the picture is **thinking or feeling.**

dispirited	ashamed	joking	confident
○	○	○	○

Continue >>

Please choose and select which word best describes what the person in the picture is **thinking or feeling.**

indecisive	nervous	ashamed	suspicious
○	○	○	○

Continue >>

Please choose and select which word best describes what the person in the picture is **thinking or feeling.**

annoyed	arrogant	terrified	upset
○	○	○	○

Continue >>

Please choose and select which word best describes what the person in the picture is **thinking or feeling.**

bored	comforting	irritated	playful
○	○	○	○

Continue >>

Please choose and select which word best describes what the person in the picture is **thinking or feeling.**

reassuring	cautious	joking	arrogant
○	○	○	○

Continue >>

Please choose and select which word best describes what the person in the picture is **thinking or feeling.**

relaxed	amused	insisting	joking
○	○	○	○

Continue >>

Please choose and select which word best describes what the person in the picture is **thinking or feeling.**

contented	joking	affectionate	interested
○	○	○	○

Continue >>

Please choose and select which word best describes what the person in the picture is **thinking or feeling.**

incredulous	panicked	interested	despondent
○	○	○	○

Continue >>

Please choose and select which word best describes what the person in the picture is **thinking or feeling.**

hostile	grateful	flirtatious	disappointed
○	○	○	○

Continue >>

Please choose and select which word best describes what the person in the picture is **thinking or feeling.**

convinced	joking	desire	flustered
○	○	○	○

Continue >>

Please choose and select which word best describes what the person in the picture is **thinking or feeling.**

alarmed	shy	hostile	anxious
○	○	○	○

Continue >>

Please choose and select which word best describes what the person in the picture is **thinking or feeling.**

terrified	flirtatious	regretful	amused
○	○	○	○

Continue >>

Please choose and select which word best describes what the person in the picture is **thinking or feeling.**

nervous	puzzled	contemplative	insisting
○	○	○	○

Continue >>

Please choose and select which word best describes what the person in the picture is **thinking or feeling.**

aghast	impatient	alarmed	fantasizing
○	○	○	○

Continue >>

Please choose and select which word best describes what the person in the picture is **thinking or feeling.**

confused	panicked	embarrassed	fantasizing
○	○	○	○

Continue >>

Please choose and select which word best describes what the person in the picture is **thinking or feeling.**

friendly	sarcastic	worried	irritated
○	○	○	○

Continue >>

Please choose and select which word best describes what the person in the picture is **thinking or feeling.**

apologetic	uneasy	dispirited	friendly
○	○	○	○

Continue >>

Please choose and select which word best describes what the person in the picture is **thinking or feeling.**

anticipating	shy	threatening	decisive
○	○	○	○

Continue >>

Big 5 inventory in Study 3 (14)

Here are a number of personality traits that may or may not apply to you.

Please write a number next to each statement to indicate the extent to which *you agree or disagree with that statement*. You should rate the extent to which the pair of traits applies to you, even if one characteristic applies more strongly than the other.

1 – Disagree strongly

2 – Disagree moderately

3 – Disagree a little

4 – Neither agree or disagree

5 – Agree a little

6 – Agree moderately

7 – Agree strongly

I *see myself as:*

1. Extraverted, enthusiastic.

2. Critical, quarrelsome.

3. Dependable, self-disciplined.

4. Anxious, easily upset.

5. Open to new experiences, complex.

6. Reserved, quiet.

7. Sympathetic, warm.

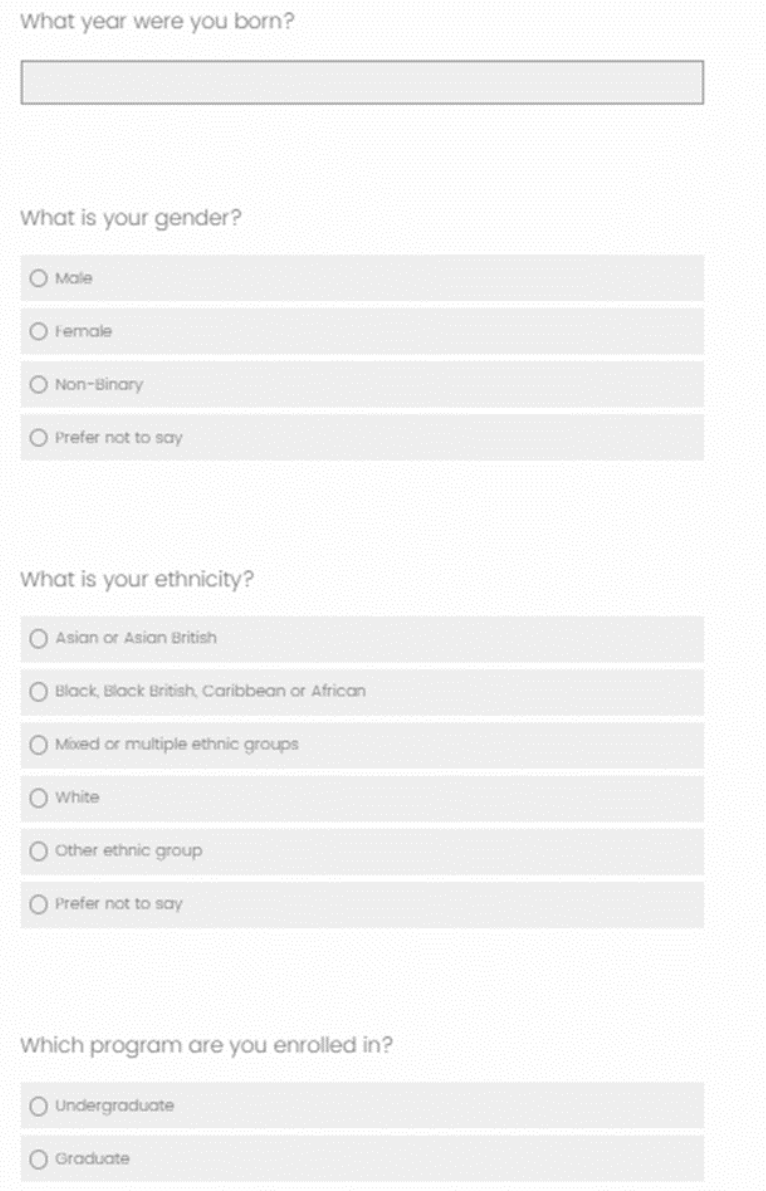
8. Disorganized, careless.

9. Calm, emotionally stable.

10. Conventional, uncreative.

Demographic questionnaire for Study 3

Note: for a full description of Study 3 materials, please see (15)



SI References

1. D. Keltner, Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame. *J. Pers. Soc. Psychol.* **68**, 441–454 (1995).

2. P. Ekman, “The Directed Facial Action Task:Emotional Responses Without Appraisal” in *Handbook of Emotion Elicitation and Assessment*, J. A. Coan, J. J. B. Allen, Eds. (Oxford University PressNew York, NY, 2007), pp. 47–53.

3. A. M. Perkins, S. L. Inchley-Mort, A. D. Pickering, P. J. Corr, A. P. Burgess, A facial expression for anxiety. *J. Pers. Soc. Psychol.* **102**, 910–924 (2012).

4. K. R. Scherer, H. Ellgring, Are facial expressions of emotion produced by categorical affect programs or dynamically driven by appraisal? *Emotion* **7**, 113–130 (2007).

5. K. R. Scherer, H. Ellgring, Multimodal expression of emotion: Affect programs or componential appraisal patterns? *Emotion* **7**, 158–171 (2007).

6. D. T. Cordaro, *et al.*, The recognition of 18 facial-bodily expressions across nine cultures. *Emotion* **20**, 1292–1300 (2020).

7. P. Rozin, A. B. Cohen, High frequency of facial expressions corresponding to confusion, concentration, and worry in an analysis of naturally occurring facial expressions of Americans. *Emotion* **3**, 68–75 (2003).

8. D. Matsumoto, P. Ekman, The Relationship Among Expressions, Labels, and Descriptions of Contempt. *J. Pers. Soc. Psychol.* **87**, 529–540 (2004).

9. D. T. Cordaro, M. Brackett, L. Glass, C. L. Anderson, Contentment: Perceived Completeness across Cultures and Traditions. *Rev. Gen. Psychol.* **20**, 221–235 (2016).

10. C. F. Benitez-Quiroz, R. B. Wilbur, A. M. Martinez, The not face: A grammaticalization of facial expressions of emotion. *Cognition* **150**, 77–84 (2016).

11. J. Reeve, The face of interest. *Motiv. Emot.* **17**, 353–375 (1993).

12. K. M. Prkachin, The consistency of facial expressions of pain: a comparison across modalities. *Pain* **51**, 297–306 (1992).

13. J. L. Tracy, R. W. Robins, Show Your Pride: Evidence for a Discrete Emotion Expression. *Psychol. Sci.* **15**, 194–197 (2004).

14. S. D. Gosling, P. J. Rentfrow, W. B. Swann, A very brief measure of the Big-Five personality domains. *J. Res. Personal.* **37**, 504–528 (2003).

15. B. Weidmann, J. Vecci, F. Said, D. J. Deming, S. R. Bhalotra, How Do You Find a Good Manager? [Preprint] (2024). Available at: https://www.nber.org/papers/w32699 [Accessed 22 July 2024].