Overview of the Emotional Recognition Task (Guess the Emotion)

Authors: Roger Strong and Amy Price

Table of Contents:

Background & Scientific Purpose

Social interactions play a critical role in many aspects of everyday life, from building new relationships to seeking new career opportunities. Inferring how others are feeling is essential to successful social interactions, as misunderstanding someone's emotional state can have negative consequences. Because people do not always verbalize their emotional state, the ability to recognize emotional facial expressions is an important part of social cognition.

The TestMyBrain Multiracial Emotion Recognition test (Deveney et al., 2018, 2022; Dodell-Feder et al., 2020; Germine et al., 2022) measures emotion recognition performance by asking participants which of four emotions (happiness, sadness, anger, fear) is expressed in images of actors' faces. The test is an adaptation of the Penn Emotion Recognition (ER-40) task (Pinkham et al., 2016), but with more diverse faces, omission of the neutral condition (which has low correlations with performance on the other conditions), and selection of images from a broader item bank to reduce ceiling effects and increase reliability. Importantly, compared to other tests of emotion recognition, the test requires minimal vocabulary knowledge and includes actors spanning a diverse range of ages and races/ethnicities. Thus, the TestMyBrain Multiracial Emotion Recognition test more specifically measures emotion recognition ability than tests where performance is strongly related to socioeconomic class and culture (Dodell-Feder et al., 2020). The test has previously been used by researchers to test emotion recognition's association with personality traits (Deveney et al., 2022) and indicators of social standing (Deveney et al., 2018, Dodell-Feder et al., 2020).

Methodology

TestMyBrain Multiracial Emotion Recognition assesses basic emotion recognition performance, using a four-alternative forced choice emotion recognition paradigm. Actors were recruited through the Boston Company One Theater, an organization dedicated to inclusivity and representation in theater, as part of the Act Out For Brain Health initiative. Importantly, although this test was designed to be demographically inclusive, it is NOT intended to permit comparisons of emotion recognition performance associated with different demographic characteristics of the actors themselves (e.g., actor race, ethnicity, age, gender-identity). Therefore, demographic information about the actors is not included. See Passell et al. (2019) for additional information about the development of this test and its psychometric characteristics.

See Figure 1 for an overview of the test structure. After reading brief instructions, participants are sequentially presented with 48 images of actors expressing either happiness, sadness, anger, or fear, and asked to identify which of the four emotions best describes the facial expression in each photo. There are no practice trials for this test. At the beginning of each test trial (see Figure 2), the participant views a single image with four response buttons beneath it: (1) Angry, (2) Fearful, (3) Sad, (4) Happy. If the participant does not make a response within 10 seconds after the test trial begins, the image is replaced by a black rectangle. If the participant does not make a response within 20 seconds after the test trial begins, a warning message appears, and participants must click a button to restart the same test trial. Participants repeat the test trial until a response is made.

Participants may select response buttons using either a mouse click or by tapping the response button on touch-compatible devices. The 48 trials are divided into 12 trials per emotion category (anger, fear, sadness, and happiness). The randomly determined trial order is the same for all participants (and for all instances of the test if participants choose to retake it after the 30-day waiting interval). The participant’s progress is visible throughout the test as a message displaying the current trial number and the total number of trials.

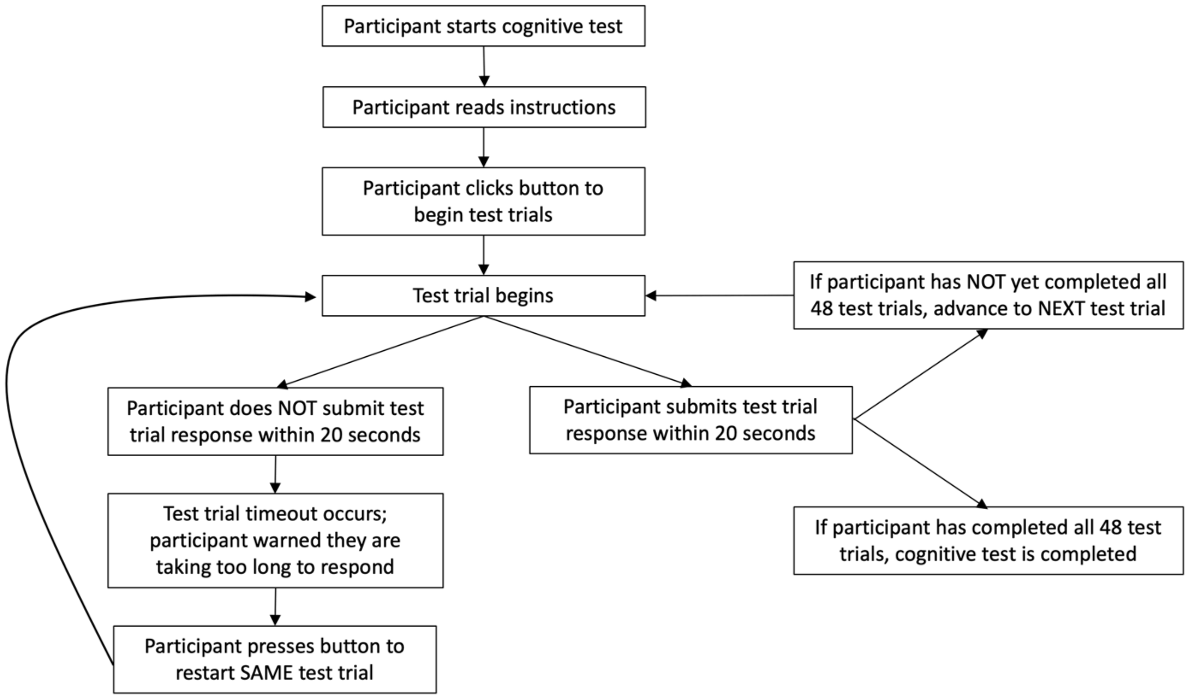

Figure 1: Flow diagram of TestMyBrain Multiracial Emotion Recognition

Figure 2: Trial structure of TestMyBrain Multiracial Emotion Recognition

Data & Analysis Guidelines

Data

As described in Introduction to Cognitive Testing Data in the *All of Us* Research Program there are three main categories of data available for cognitive tests: (1) trial-level data, (2) summary scores, and (3) metadata. Please see the Exploring the Mind Data Dictionary here.

Suggested Outcomes

The test's suggested primary outcome is *accuracy*: the proportion of trials answered correctly. Higher accuracy indicates better emotion recognition performance. As secondary outcomes, researchers might consider separately analyzing performance for the four different emotions that are displayed throughout the test (*happy_accuracy, sad_accuracy, fearful_accuracy, angry_accuracy)*. Researchers might also consider analyzing variables related to response speed (e.g., *medianRTc*) and response time variability (e.g., *sdRTc*).

Outcome Type	Outcome Name	Description
Primary	accuracy	Proportion of trials answered correctly. Higher accuracy indicates better emotion recognition performance
Secondary	happy_accuracy	Proportion of trials answered correctly where the correct answer was "happy"
	sad_accuracy	Proportion of trials answered correctly where the correct answer was "sad"
	fearful_accuracy	Proportion of trials answered correctly where the correct answer was "fearful"
	angry_accuracy	Proportion of trials answered correctly where the correct answer was "angry"
	medianRTc	Median reaction time of correct responses
	sdRTc	Standard deviation of reaction times of correct responses

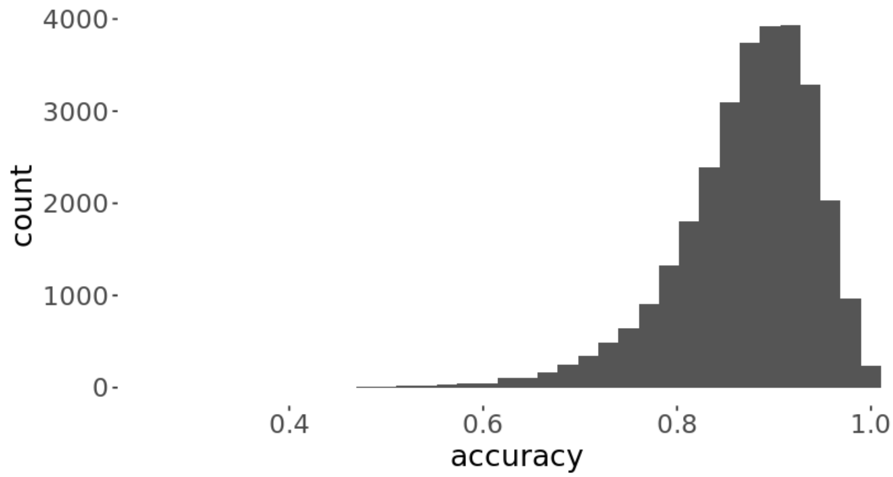

Figure 3: Histogram of primary outcome metric (accuracy) for all participants in the Curated Data Repository (CDR) v8 off-cycle release

Quality Control Guidelines

The following guidelines are provided for the purpose of flagging extreme deviations in performance from what is typically seen in participants performing the task in a valid manner. Researchers must use their own judgment when determining whether flagged participants should be excluded from analyses. Researchers may also consider implementing their own quality control criteria separately from these recommendations. For more details about quality control criteria, please see Introduction to Cognitive Testing Data in the *All of Us* Research Program.

Quality control variables are provided both in trial-level data and full-test summary data. The table below summarizes the quality control variables available for this test.

Flag Type	Variable Name	Description
Trial-level	flagged	Indicates whether a participant's reaction time (rt) for a trial is less than 300 ms (1 if rt < 300, 0 if rt >= 300, null if no response). Accurately responding in under 300 ms on manual choice reaction time tasks is physiologically implausible.
Full-test	flag_medianRTc	Has a value of 1 when a participant's median response time for correctly answered trials (medianRTc) is less than 700 ms. The accuracy of prior participants completing this test was distributed around chance-level (.25) when medianRTc was less than 700

		ms, suggesting these participants were randomly guessing to complete the test quickly.
	flag_sameResponse	Has a value of 1 when participants select the same response option (fearful, angry, happy, or sad) on more than 90% of the trials, and a value of 0 otherwise. Participants selecting the same answer repeatedly are likely not engaged with the test.
	flag_trialFlags	Has a value of 1 when more than 10% of a participant's trials are flagged for having reaction time less than 300 ms (flagged=1), and a value of 0 otherwise. The accuracy of prior participants completing this test was distributed around chance-level (.25) when more than 10% of trials were answered in less than 300 ms, suggesting these participants were randomly guessing to complete the test quickly.

Emotion Recognition Task (N = 29,948)[1]		
	Yes	**No**
Median RT Flags	<1%[2]	>99%
Same Response Flags	<1%	>99%
Trial Flags	<1%	>99%
Any Flags	<1%	>99%

Table 1: Percentage of participants with quality control flags in the Exploring the Mind CDR v8 off-cycle release.

Calculating Test Reliability

To calculate the reliability of Multiracial Emotion Recognition performance differences between participants in a given sample, we recommend using one of the following two approaches:

1. Compute Cronbach's alpha (Cronbach, 1951; Tavakol & Dennick, 2011), using the accuracy of each participant's responses (1/0) to each of the 48 trials.
2. Calculate *split-half reliability* (Pronk et al., 2022) using the following steps:

[1]This count is defined as the total number of unique participants who completed the task.
[2]Due to the data dissemination policy, counts of less than 20 participants cannot be shared publicly. Users can view exact counts in the corresponding featured workspace after logging into their Researcher Workbench account.

2. For each participant, compute accuracy separately for odd trials (trials 1, 3, 5, etc.) and even trials (trials 2, 4, 6, etc.).
3. Compute the Pearson correlation (r) between (1) accuracy on odd trials and (2) accuracy on even trials.
4. Use the Spearman-Brown prediction formula to compute full-test reliability: reliability = (2*r) / (1+r)

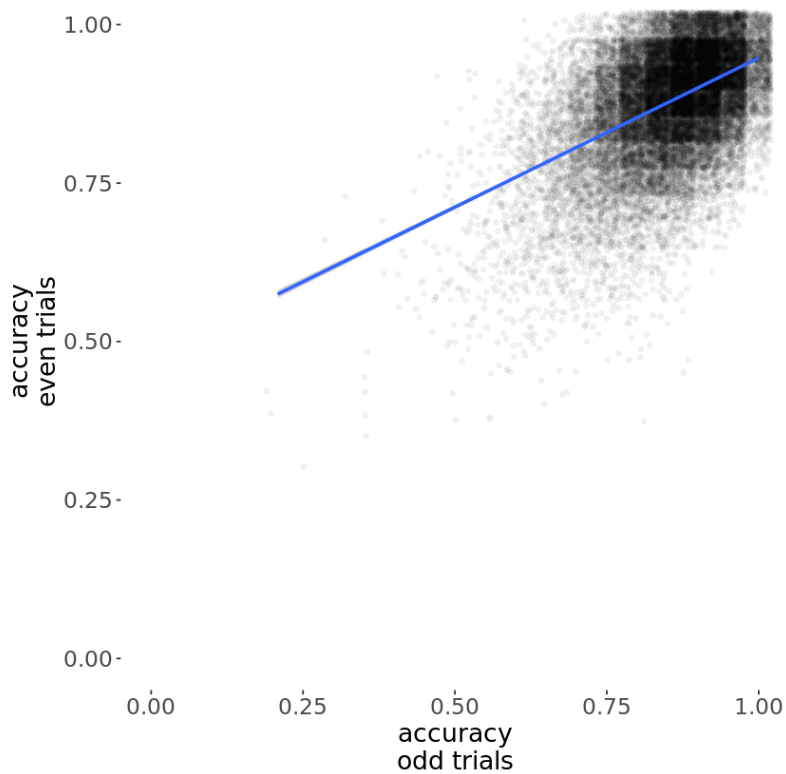

Figure 4: Correlation of participants' Multiracial Emotion Recognition accuracy on even and odd trials (Spearman-Brown split-half reliability = 0.65)

-

Correlates of Interest

Prior data collection has found associations between the following demographic variables and performance (*accuracy*) on TestMyBrain Multiracial Emotion Recognition. Therefore, researchers may consider including the following variables as covariates in analyses.

1. *age:* prior reports have found associations between age and accuracy, although the precise relationship differed between reports, likely due to differences in statistical approaches used. One report found that older age was associated with better accuracy until around age 18, but that after age 18 there was not an association between age and performance (Dodell-Feder et al., 2020). A second report found that older age was

associated with better accuracy until around age 35, after which older age was associated with worse performance (Passell et al., 2019)

2. *Biological sex assigned at birth*[3]: reporting sex as *female* is associated with a small accuracy advantage relative to reporting sex as *male* (Dodell-Feder et al., 2020). Researchers may also consider including sex assigned at birth as a covariate in analyses.
3. *education*: educational attainment has a small association with better accuracy (Deveney et al. 2018; Dodell-Feder et al., 2020).

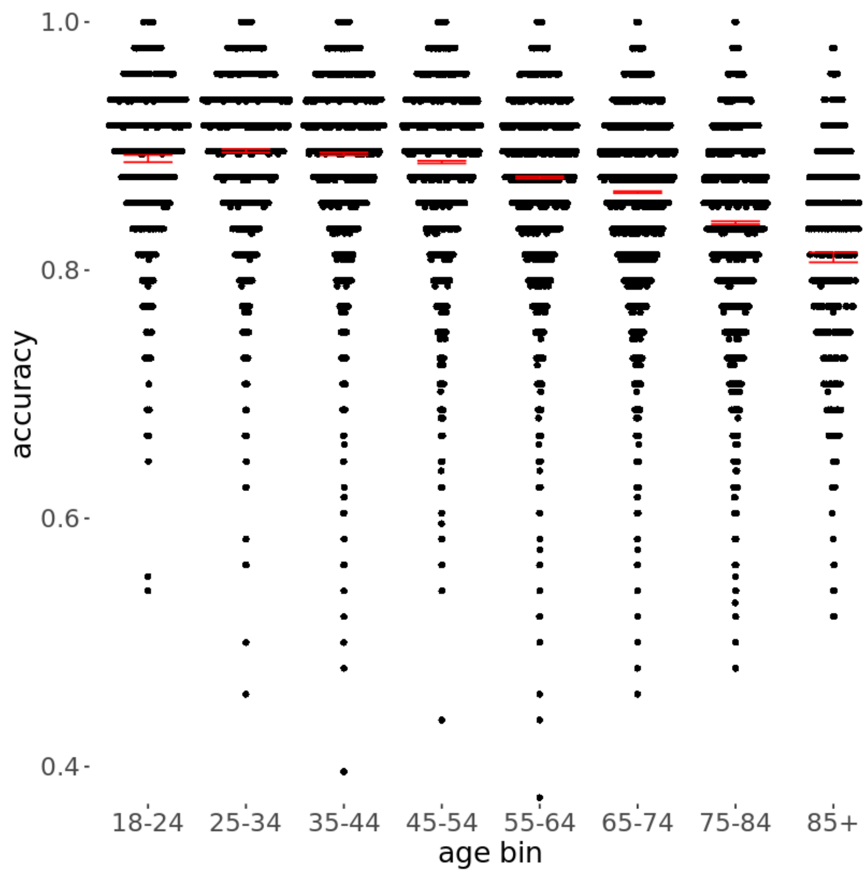

Figure 5: Accuracy by age bin. Red lines represent mean accuracy for each bin. Width of distributions (black dots) represent the relative density of participants at each accuracy level.

[3]Sometimes referred to as gender in prior research publications

References

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334. https://doi.org/10.1007/BF02310555

Deveney, C. M., Chavez, G., & Mejia, L. (2022). Trait irritability in adults is unrelated to face emotion identification. *Personality and Individual Differences, 185*, 111290. https://doi.org/10.1016/j.paid.2021.111290

Deveney, C. M., Chen, S. H., Wilmer, J. B., Zhao, V., Schmidt, H. B., & Germine, L. (2018). How generalizable is the inverse relationship between social class and emotion perception?. *PloS One*, *13*(10), e0205949. https://doi.org/10.1371/journal.pone.0205949

Dodell-Feder, D., Ressler, K. J., & Germine, L. T. (2020). Social cognition or social class and culture? On the interpretation of differences in social cognitive performance. *Psychological Medicine*, *50*(1), 133-145. https://doi.org/10.1016/j.psychres.2019.112682

Germine, L. T., Joormann, J., Passell, E., Rutter, L. A., Scheuer, L., Martini, P., ... & Kessler, R. C. (2022). Neurocognition after motor vehicle collision and adverse post-traumatic neuropsychiatric sequelae within 8 weeks: Initial findings from the AURORA study. *Journal of Affective Disorders, 298*, 57-67. https://doi.org/10.1016/j.jad.2021.10.104

Passell, E., Dillon, D. G., Baker, J. T., Vogel, S. C., Scheuer, L. S., Mirin, N. L., … Germine, L. (2019, January 26). Digital Cognitive Assessment: Results from the TestMyBrain NIMH Research Domain Criteria (RDoC) Field Test Battery Report. https://doi.org/10.31234/osf.io/dcszr

Pinkham, A. E., Penn, D. L., Green, M. F., & Harvey, P. D. (2016). Social cognition psychometric evaluation: Results of the initial psychometric study. *Schizophrenia Bulletin, 42*(2), 494-504. https://doi.org/10.1093/schbul/sbv056

Pronk, T., Molenaar, D., Wiers, R. W., & Murre, J. (2022). Methods to split cognitive task data for estimating split-half reliability: A comprehensive review and systematic assessment. *Psychonomic Bulletin & Review, 29*(1), 44-54. https://doi.org/10.3758/s13423-021-01948-3

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education, 2*, 53-55. https://doi.org/10.5116%2Fijme.4dfb.8dfd