Quicker, extremer: a computational modeling of rating and reaction time in social evaluation of faces

Nan Wang (nanwanglin@Link.Cuhk.Edu.Hk)
Department of Linguistics, Central Avenue
Sha Tin, 999077 Hong Kong

Abstract

When individuals are pressed to make decisions quickly, their accuracy tends to decline, which is termed as speed-accuracy trade-off. But does this phenomenon extend to perceptual rating? In other words, do rapid judgments result in more extreme outcomes? To address this question, the study analyzed a global dataset covering 11,481 adult participants' ratings of 120 targets across 45 countries. The hypothesis posited that the rating became more extreme if it took less time. The study firstly identified response time as a significant predictor in extremity of social judgments through a machine learning algorithm, XGBoost, with cultural variables emerging as the second most important predictor. Given the importance of response time, the study employed hierarchical general linear models to investigate whether faster decision-making correlates with more extreme ratings and how this effect varies across diverse cultural contexts. The findings revealed a significant global level effect, also showing considerable variance across eleven regions. This observed phenomenon is termed as the "speed-extremity trade-off," and is strongest in the Middle East and weakest in East Southeast Asia and Scandinavia.

Keywords: perceptual decision making, machine learning

Introduction

People often form an impression of others based on relevant facial cues in a short period. Extremely brief exposures as short as 140ms are shown to be enough for people to make social judgments of faces (Crouzet et al., 2010). And a ton of studies have been conducted to study the information about identity, mental, and emotional states conveyed by human faces (Todorov et al., 2013). However, in revealing the mechanisms of such a rapid process of facial perception, studies are mostly limited to two areas: 1) how people integrate information of facial attributes such as race, gender, and 2) how external factors such as culture-related variables, region and country, interact with the process (Tipples, 2023). In the studies above, time is usually seen as an irrelevant variable and is seldom taken into consideration in data analysis. Two threads of recent evidence, however, challenges this conception, and suggests that response time can also contain rich information.

Firstly, across paradigms or stimuli, response time(RT) can have different distribution patterns, providing implicit but valuable insights into what is being rated and how faces are rated. In face categorization tasks, researcher found that, for decisions related to familiar and unfamiliar faces, there are temporal shifts in the shape of RT distributions, while the distribution of response time to self faces remained stable, irrespective of the context and task demands(Sui et al., 2013). Relatedly, a study on self-advantage in face recognition directly focused on response time as the primary object of investigation(Y. Ma & Han, 2010). Some other paradigms manipulate response time to reveal the processes and related effects in facial judgments. In a recent study, research found that with the increase of face presentation time, the influence of facial appearance on facial attractiveness increases. Furthermore, as response deadlines increase, participants are more likely to associate facial attractiveness with moral behavior (Li et al., 2023). This suggests that the duration of response time may impact participants' performance in facial judgments.

Secondly, in the temporal dimension of facial judgments, studies have also found that facial decision-making exhibits different characteristics at different time points during the process(Dobs et al., 2019). The study revealed that during the process of detecting and recognizing faces, different response time points indicate distinct stages of individual face information processing. For example, based on the analysis of response time, study has found that the brain encodes gender and age information before identity information, and early processing of gender and identity information is enhanced based on the familiarity of the faces. This suggests that response time can reflect the stage of cognitive processes.

Given all the information contained in response time, the central question we are interested in is how response time itself shapes ratings. Specifically, this study focuses on the aspect of extremity in ratings for the following reasons.

In previous literature, the relation between reaction time and accuracy has been explored in a phenomenon called speed accuracy tradeoff, where individuals must balance the need to respond quickly with the need to provide an accurate answer when making judgments(MacKay, 1982; Plamondon & Alimi, 1997). In the context of perceptual decision-making such as facial ratings where we don't have right or wrong standards, the accuracy can be translated as extremity where we treat extreme ratings that fall at the two ends of a scale as "inaccurate" ratings. This interpretation is both relevant and appropriate for several reasons. Firstly, extreme ratings, whether highly pleasant or unpleasant, are less common occurrences. Secondly, the dataset used for facial stimuli, sourced from Jones (Jones, 2021), comprises neutral faces. Thus, it's reasonable to classify highly extreme ratings as "inaccurate".

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

Thus, answers to the central question, how does response time itself correlates with ratings, is guided by three considerations: 1) how important response time is, when taken as an independent variable, in comparison with other variables in social judgment of faces; 2) how does this response time influence extremity of ratings in general, 3) how does this effect vary across races or cultural contexts?

Results

The importance of response time

The first question we want to answer is how important response time is, when compared with other variables?

To answer the question, we firstly divided the variables into different groups. Table 1 presents the descriptive statistics of variables used in this analysis. The extremity of ratings (extreme or non-extreme) is identified as target variables. Response time, stimulus-related, and culture-related variables are selected features that have been previously shown to influence extremity in ratings(Batres & Shiramizu, 2023; Hester et al., 2021). Among them, stimulus-related variables include race, sex, gender of the stimulus. As to culture-related variables, language is chosen as previous studies have shown that it can shapes cognitive processes and influences the interpretation of facial expressions and traits(Landau et al., 2010; Lindquist & Gendron, 2013; Tsao & Livingstone, 2008) . Region is taken as another metric of culture-related variable. Culture is operationalized as regions in another study as well(Hester et al., 2021).

Table 1. A list of variables

Category	Variables		Levels
Features	Extremity of ratings		Extreme, moderate
Labels	Stimulus related	race	Black, Latin, Asian, White
		Gender	Male, female
		Age	Continuous
	Culture related	Region	Africa, East and Southeast Asia, Australia and New Zealand, Middle East, etc. (11 regions)
		Language	English, French, etc. (25 languages)
	Response time		Continuous

To investigate the multiple factors that contribute to extreme ratings, we used a machine learning method, XGBoost to predict the importance of each (Chen & Guestrin, 2016) . In contrast with traditional classification methods such as random forest, XGBoost creates trees sequentially instead of parallelly. This ensemble method builds models in a sequential manner and can combine several weak learners into a strong one to improve accuracy of predicting (Géron, 2017). The XGBoost model is trained using response time, stimulus-related factors, and culture-related factors as input variables.

We first split the data into a training set and a test set, with 70% of the data used for training and 30% for testing. The trained model was then used to make predictions on the test data, and the performance was evaluated using a confusion matrix, which is presented in Figure x. The rows represent the actual class labels, while the columns represent the predicted class labels. The elements of the table display the counts of data points falling into each combination of actual and predicted classes. And the accuracy of the model is as high as 73.58%.

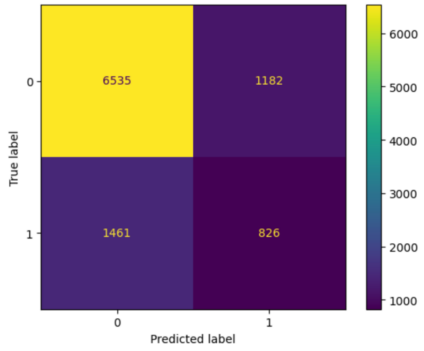

Figure 1: Confusion matrix

A graph was generated to visualize decision trees and their corresponding splits. In this graph, each node represents a feature, with the most influential one sitting at the top, which is response time. Then, each branch represents a decision outcome and a threshold value based on that feature, and each leaf represents the prediction.

Figure 2: Decision tree of XGBoost model

The feature importance results are also plotted using a radar plot, showing the relative significance of predictor variables. The features are positioned along the axes of the radar plot, with each axis representing a specific predictor variable. The radial distance from the center of the plot to each feature's point illustrates its corresponding importance score. The radar plot reveals that response time holds the highest importance, followed by language_NL (Dutch), region_Africa (indicating whether the region is Africa), and race (indicating whether the race is Latino). That is to say, besides response time, culture-related variables are the second most important variables in predicting extremity of ratings.

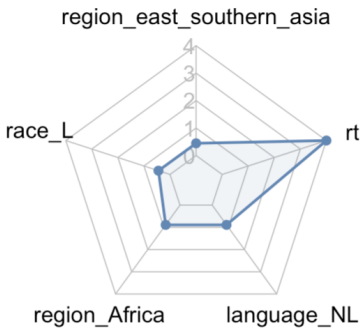

Figure 3: The five most important features

Examine speed-extremity trade-off in global and culture-specific context

Now that we have identified the most important variable, the next step seems to explore the relations between response time and ratings. However, there's a lot of heterogeneity hidden in the dataset, as response time and ratings are nested within higher-level groups, such as region, language, race, etc. Therefore, we need to find the group variable that has smallest within-group variance. And to achieve, we used intercept models of hierarchical linear modeling (HLM), consisting of two levels: an individual rating level (Level 1) that represents the within-group variation and includes observation-specific predictors and the intercept term; a group level (Level 2) that explains the variability between different groups. Table 2 displays the Intraclass Correlation Coefficients (ICCs) for predicting reaction time using different group variables. Notably, "user_id" exhibits a substantial ICC of 0.345, indicating significant variability. It's essential to clarify that "user_id" pertains to individual participants, each contributing 120 ratings (equivalent to 120 rows of observations).

Table 2. The ICC of different group variables

Factors	ICC
User id	0.345
Race	0.001
Trait	0.052
Language	0.012
Country	0.057

Now that we have known response time is important and user ID is the proper group variable, the next step is to look at how response time and rating extremity positively or negatively correlates with each other and how does the effect vary across different cultural contexts. Importantly, the study operationalizes culture as eleven regions, consistent with previous practices (Hester et al., 2021). For a specification of the world regions and corresponding countries, see Methods section. The study built a HLM model for each region respectively. And the slope of each model is taken, representing the direction and magnitude of the association between quicker response times and more extreme ratings. All the values are negative, suggesting a negative relationship between the two: quicker judgments correlate with more extreme ratings. Also, the most robust effect is observed in the Middle East (0.254). Conversely, East and Southeast Asia, which covers China, India, Malaysia, Taiwan, Thailand, exhibits a relatively weaker effect (0.114).

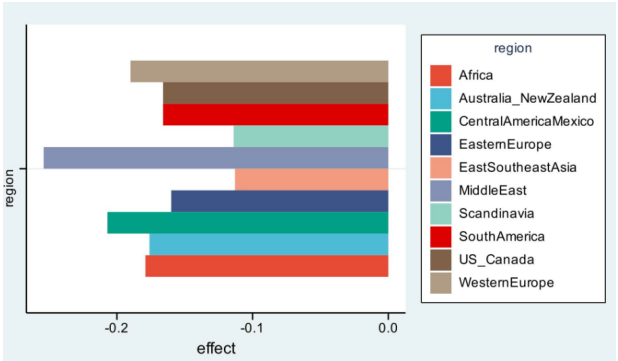

Figure 4: The strength of effect across regions

Discussion

Although there're already many studies investigating the interaction between response time and accuracy, especially in the two-choice perceptual decision-making tasks, our study is one of the first to see whether they are applicable in a face evaluation setting, where 1) multiple choices, that is, a range of ratings on a scale, are given, 2) there's no right or wrong answers. This study extends previous conclusions and models in speed-accuracy trade-off to a facial perception setting. And the major findings are summarized and interpreted below.

Firstly, response times turns out to be the most important variables in predicting the extremity of ratings, and is seconded by region and languages, which belong to the category of culture-related variables. This means that culture-related variables are important in shaping people's perception and judgment of a certain trait. This means that the culture background, conservative, progressive, etc., can also have an impact on people's tendency of giving extreme ratings. This finding also allies with previous studies revealing culture modulates the perception of faces(Keating et al., 1981; Masuda et al., 2008; Voegeli et al., 2021).

Secondly, the within-group variance is smallest when grouped under a single participant who needs to complete 120 ratings. This suggests there're a lot of individual differences even in perceptual decision-making domain, making them a crucial factor in understanding facial evaluations.

Thirdly, response time negatively predicted extremity of ratings, which means that when people make decisions more quickly, their ratings tend to be more extreme, either towards the high end or the low end of the scale. This extends the

speed-accuracy trade-off in cognitive decision-making to perceptual decision-making: the decisions can be not only less accurate, but also more biased (i.e. more extreme), when they are made in a shorter period. But why is it the case? One explanation is that quick judgments often rely on mental shortcuts or heuristics, which can introduce biases(Gilovich et al., 2002). These biases may lead to more extreme conclusions if the judgment is influenced by emotional reactions or cognitive shortcuts that tend to amplify certain aspects of a situation.

Also, the effect, which we termed as speed-extremity trade-off is not uniform globally; there are regional differences. The most prominent speed-extremity trade-off is observed in the Middle East. In comparison, the effect is least pronounced in Scandinavia and East and Southeast Asia. One plausible explanation for the least pronounced effect in East Southeast Asia lies in the cultural attributes: as the society values consensus-building and collective decision-making, individuals might be less inclined to make rapid, extreme judgments(Monkhouse et al., 2013). In contrast, in the Middle East, the pronounced speed-extremity trade-off can be a result of long-time geopolitical influences. Its history, marked by conflicts and geopolitical tensions, may have fostered a heightened sense of urgency and the need for quick decision-making in certain situations(Salloukh, 2013). The relatively pronounced effect in Western Europe can instead be explained by fast-paced nature of modern life, pressure to meet deadlines, and a desire to demonstrate competence and productivity.

Conclusion

In summary, the speed at which individuals made a rating is closely related with the results of their rating and we termed it as a speed-extremity trade-off. In regions including Middle East, Central America, Western Europe, it's more likely for an individual to increase his or her response speed at the cost of being more biased, while in Scandinavia and East and Southeast Asia, people tend to be cautious in giving extreme ratings. This difference can be attributed to culture norm differences: in cultures that value collectivism, harmony, and caution, such as parts of East and Southeast Asia and Scandinavia, there may be a preference for more deliberate and cautious decision-making processes; while in Middle East, the long-time geopolitical tensions can cause people to react in a quicker and extremer way.

In the future, more nuanced models can be built to understand the accumulation process over the course of this type of effect. For example, random walk models, one of the oldest models built to model choice-reaction time relation (Stone, 1960), can be used. And mixture models, proposed by Ollman (1966), instead identifies two major components in the process: fast guesses and slow controlled decisions. Both models can be leveraged to reveal the underlying mechanism of this effect.

Methods

Sample

The sample used by this study comes from a global dataset collected by Jones (2021). This dataset covers 11,481 adult participants' ratings of 120 targets across 45 countries or regions. And the 45 five countries are further divided into 11 regions.

In Jones' study, the stimuli come from the Chicago face dataset(Ma et al., 2015) and consist of the faces of 60 men and 60 women and are equally divided into four races: Black, Asian, Black, White, and Latino. Participants rated faces for 14 traits on an ordinal scale from 1 to 9, including aggressive, attractive, caring, confident, dominant, emotionally stable, intelligent, mean, responsible, sociable, trustworthy, unhappy, or weird. Each rater was randomly assigned a certain trait to evaluate for all 120 faces. Besides, the demographic information of participants, including sex, age, and ethnicity, is also collected through questionnaires.

We firstly conduct exploratory data analysis to see how important response time is to extremity of ratings, when compared with other variables. To be more specific, we applied the XGBoost machine learning algorithm to assess the importance of response time compared to stimulus-related or culture-related factors in predicting extremity of ratings.

After that, based on the structure of the data, we use the Hierarchical General Linear Model approach to capture variability of the effect of reaction time on rating extremity. Specifically, in the first level of the model, we use scoring extremity as the dependent variable and reaction time as the predictor variable, in the second level of the model we put the user id , and in the third level of the model we put traits.

Eventually, we look at the global relationship between response time and ratings using the same generalized linear model. We also include culture-related variables identified previously into the model to examine their effects on ratings.

Preprocessing

The original dataset has 682, 545 observations. Given the limited computing power we have, we only randomly sampled one part of this original dataset for our analysis which finally gives 95, 423.

Firstly, we excluded raters that didn't complete 120 ratings or gave same ratings for 75% or more faces. Then data quality will also be checked by calculating Cronbach's α and test–retest reliability, following the same criteria of Jones (2021).

Then we applied a log transform to the response time data to stabilize the variance and mitigate the impact of extreme values. And outliers are eliminated from the dataset by removing any RT values falling beyond three standard deviations from the mean.

To capture the deviation of each rating from the average, the mean value (five) was subtracted from each rating and the absolute value of the subtracted ratings was taken. In this way, the original rating data was effectively rescaled to a new

range that encompasses a spectrum from non-extreme to the most extreme rating.

And to split ratings into extreme and non-extreme categories, we define extreme ratings as those falling at the endpoints of the ordinal scale (1 or 9), and non-extreme ratings as those falling closer to the midpoint (4 or 6). This can create a clear distinction between the two categories and allow for a more straightforward interpretation of the data. And instead of selecting ratings of 5 as moderate ratings, we choose 4 or 6 because it can help mitigate potential response biases that would occur when participants feel uncertain or impatient and just jump to middle-of-the-road ratings (e.g., 5). Also, considering the sample size, defining a broader range as non-extreme (e.g., 4 or 6) may increase the likelihood of capturing more extreme responses and allow for more robust statistical comparisons between extreme and non-extreme ratings.

XGBoosting

Model Our dataset $\{(\boldsymbol{x_i},\ \boldsymbol{y_i})\}_{\boldsymbol{i=1}}^{\boldsymbol{n}}\ (\boldsymbol{x_i} \in \boldsymbol{R^m}, \boldsymbol{y_i} \in \boldsymbol{R})$ has 95423 observations, and 100 trees in total, and $\boldsymbol{x}$ stands for the combination of predictors while y stands for observed variable of extremity. Our goal is to minimize the following regularized objective function(Chen & Guestrin, 2016):

$$L(\emptyset) = \sum_i L(\hat{y}_i, y_i) + \sum_k \Omega(f_k)$$

$$\text{where } \Omega(f) = \gamma T + \frac{1}{2}\lambda\|w\|^2$$

Ω penalizes the complexity model to prevent overfitting, and λ represents the regularization parameter.

Split dataset We first split the data into a training set and a test set, with 70% of the data used for training and 30% for testing. Categorical variables, including "Race", "Gender", "language", and "region", were subjected to one-hot encoding that transforms them into binary vectors, creating new binary columns for each unique category.

Training and evaluation Because XGBoost tend to overfit by fitting complex decision boundaries, we used a grid search with cross-validation to tune the hyperparameters, especially regularization hyperparameters to prevent overfitting. The grid search was performed using the GridSearchCV function from the scikit-learn library, with 3-fold cross-validation. The objective function was set to 'binary:logistic', and a random seed of 42 was used.

- Max depth: [3, 4, 5]
- learning rate: [0.1, 0.01, 0.05]
- gamma: [0, 0.25, 1.0]
- reg lambda: [0, 1.0, 10.0]

The best combination of hyperparameters is determined based on lowest root mean squared error (RMSE): max_depth = 4, learning_rate = 0.1, gamma = 0.25, and reg_lambda= 1.0. These parameters were then used to initialize XGboost classifier. The model was trained on the training data using the fit function.

Feature importance analysis The feature importance scores were computed. The importance scores were calculated based on the 'weight' metric, which represents the number of times a feature was used to split the data across all trees in the model.

Hierarchical General Linear Modeling

Data were analyzed with lme4 package (Bates et al., 2015) in R by using random intercept and slopes. The dependent variable was the extreme of ratings, and we estimated the effect of reaction time on the extremity of the score with five two-level hierarchical linear models. The interclass correlation coefficients (ICC) are calculated at the same time to estimate the proportion of variance in the dependent variables that can be attributed to users, race of the stimulus, trait, languages and country of the participants. After finding the best group variable, we used HLM to investigate the relationship between extremity of rating and rt. The Model structure is given below:

$$y \sim x + (1 + x \mid \text{variable})$$

Here, y represents the extreme of ratings, x represents the reaction time of participants. And the "variable" is replaced by five different variables: users, race of the stimulus, trait, languages and country of the participants.

To investigate the variability of the observed effect across different cultural contexts, we built an HLM individually for each region, using the same model structure. Subsequently, the region-specific slopes derived from the eleven models were aggregated to illustrate the diversity in the strength of the effect across various cultural contexts. And Table 3 gives the specific countries within each region.

Table 3. Culture operationalized as eleven regions

Region	Country
Africa	Kenya, Nigeria, South Africa
East Southeast Asia	China, India, Malaysia, Taiwan, Thailand
Australia New Zealand	Australia, New Zealand
Central America Mexico	El Salvador, Mexico
Eastern Europe	Hungary, Lithuania, Poland, Russia, Serbia, Slovakia
Middle East	Iran, Israel, Turkey
US Canada	Canada, United States
Scandinavia	Denmark, Finland, Norway, Sweden
South America	Argentina, Brazil, Chile, Colombia, Ecuador
United Kingdom	England, Scotland, Wales

Western Europe	Austria, Belgium, France, Germany, Greece, Italy, The Netherlands, Portugal, Spain, Switzerland

Acknowledgement

I would like to extend my sincere appreciation to Dr. Hu Chuan-Peng who provided the inspiration for the idea of the paper. Also, I would like to thank Bai Songshi who has aided a lot in data processing. Their support has significantly contributed to this work.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software, 67(1). https://doi.org/10.18637/jss.v067.i01

Batres, C., & Shiramizu, V. (2023). Examining the "attractiveness halo effect" across cultures. *Current Psychology*, *42*(29), 25515–25519. https://doi.org/10.1007/s12144-022-03575-0

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades toward faces: Face detection in just 100 ms. *Journal of Vision*, *10*(4), 16.1-17. https://doi.org/10.1167/10.4.16

Dobs, K., Isik, L., Pantazis, D., & Kanwisher, N. (2019). How face perception unfolds over time. *Nature Communications*, *10*(1), Article 1. https://doi.org/10.1038/s41467-019-09239-1

Géron, A. (2017). *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems* (1st edition). O'Reilly Media.

Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge University Press.

Hester, N., Xie, S. Y., & Hehman, E. (2021). Little between-region and between-country variance when people form impressions of others. *Psychological Science*, *32*(12), 1907–1917. https://doi.org/10.1177/09567976211019950

Jones, B. C. (2021). To which world regions does the valence–dominance model of social perception apply? *Nature Human Behaviour*, *5*, 13.

Keating, C. F., Mazur, A., Segall, M. H., Cysneiros, P. G., Kilbride, J. E., Leahy, P., Divale, W. T., Komin, S., Thurman, B., & Wirsing, R. (1981). Culture and the perception of social dominance from facial expression. *Journal of Personality and Social Psychology*, *40*(4), 615–626. https://doi.org/10.1037/0022-3514.40.4.615

Landau, A. N., Aziz-Zadeh, L., & Ivry, R. B. (2010). The Influence of Language on Perception: Listening to Sentences about Faces Affects the Perception of Faces. *Journal of Neuroscience*, *30*(45), 15254–15261. https://doi.org/10.1523/JNEUROSCI.2046-10.2010

Lindquist, K. A., & Gendron, M. (2013). What's in a Word? Language Constructs Emotion Perception. *Emotion Review*, *5*(1), 66–71. https://doi.org/10.1177/1754073912451351

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

Ma, Y., & Han, S. (2010). Why we respond faster to the self than to others? An implicit positive association theory of self-advantage during implicit face recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(3), 619–633. https://doi.org/10.1037/a0015797

MacKay, D. G. (1982). The problems of flexibility, fluency, and speed–accuracy trade-off in skilled behavior. *Psychological Review*, *89*(5), 483–506. https://doi.org/10.1037/0033-295X.89.5.483

Masuda, T., Ellsworth, P. C., Mesquita, B., Leu, J., Tanida, S., & Van de Veerdonk, E. (2008). Placing the face in context: Cultural differences in the perception of facial emotion. *Journal of Personality and Social Psychology*, *94*(3), 365–381. https://doi.org/10.1037/0022-3514.94.3.365

Monkhouse, L. L., Barnes, B. R., & Hanh Pham, T. S. (2013). Measuring Confucian values among East Asian consumers: A four country study. *Asia Pacific Business Review*, *19*(3), 320–336. https://doi.org/10.1080/13602381.2012.732388

Ollman, R. (1966). Fast guesses in choice reaction time. *Psychonomic Science*, *6*(4), 155–156. https://doi.org/10.3758/BF03328004

Plamondon, R., & Alimi, A. M. (1997). Speed/accuracy trade-offs in target-directed movements. *Behavioral and Brain Sciences*, *20*(2), 279–303. https://doi.org/10.1017/S0140525X97001441

Salloukh, B. F. (2013). The Arab Uprisings and the Geopolitics of the Middle East. *The International Spectator*, *48*(2), 32–46. https://doi.org/10.1080/03932729.2013.787830

SALTHOUSE, T. A. (1979). Adult age and the speed-accuracy trade-off. *Ergonomics*, *22*(7), 811–821. https://doi.org/10.1080/00140137908924659

Stone, M. (1960). Models for choice-reaction time. *Psychometrika*, *25*(3), 251–260. https://doi.org/10.1007/BF02289729

Sui, J., Hong, Y., Hong Liu, C., Humphreys, G. W., & Han, S. (2013). Dynamic cultural modulation of neural responses to one's own and friend's faces. *Social Cognitive and Affective Neuroscience*, *8*(3), 326–332. https://doi.org/10.1093/scan/nss001

Tipples, J. (2023). Analyzing facial expression decision times: Reaction time distribution matters. *Emotion*

(Washington, D.C.), *23*(3), 688–707. https://doi.org/10.1037/emo0001098

Todorov, A., Mende-Siedlecki, P., & Dotsch, R. (2013). Social judgments from faces. *Current Opinion in Neurobiology*, *23*(3), 373–380. https://doi.org/10.1016/j.conb.2012.12.010

Tsao, D. Y., & Livingstone, M. S. (2008). Mechanisms of Face Perception. *Annual Review of Neuroscience*, *31*(1), 411–437. https://doi.org/10.1146/annurev.neuro.30.051606.094238

Voegeli, R., Schoop, R., Prestat-Marquis, E., Rawlings, A. V., Shackelford, T. K., & Fink, B. (2021). Cross-cultural perception of female facial appearance: A multi-ethnic and multi-centre study. *PLOS ONE*, *16*(1), e0245998. https://doi.org/10.1371/journal.pone.0245998