Manuscript

When Men are Wonderful: A Larger Happy Face Facilitation Effect for Male (vs Female) Faces for Male Participants

Jason Tipples*[1]

[1] Psychology Group, Leeds Beckett University

Keywords: Facial Expressions, Stereotypes, Evaluative Associations, Social Categorization

Author Note

Dr Jason Tipples https://orcid.org/0000-0002-0501-2129

Author Note: Data, Code and pre-registration: https://osf.io/p35v7/ .

Correspondence: Jason Tipples, Psychology Group, School of Social, Psychological & Communication Sciences Leeds Beckett University, [CL 815], City Campus, Leeds, LS1 3HE, UK.

Tel: +44 (0)113 812 23002 | Email: W.Tipples@leedsbeckett.ac.uk

Abstract

Facial cues for age, race, and sex influence how we recognize facial expressions. For example, the faster recognition of happy compared to sad expression increases in magnitude when the faces are female compared to male (Bijlstra et al., 2010; Hugenberg & Sczesny, 2006) – an effect termed Researchers (Bijlstra et al., 2010) have argued that presenting expressions of opposite valence (e.g., sad vs happy expressions) creates an evaluative mindset and consequently, face sex affects emotion recognition via evaluative rather than stereotype associations. For the comparison between anger and happiness recent results (Tipples, 2022a, 2022b) indicate that effects of face sex are larger for female participants. However, for the critical comparison between sad and happy expressions — used to support the evaluative over the stereotype account — moderation by participant sex has not been adequately examined because the sample size of male participants has been too small. Here, I increased the number of male participants relative to previous studies. For male participants, the usual facilitation effect for female faces was reversed — the happy face facilitation effect was larger for male compared to female faces. The novel pattern for male participants – supporting an ingroup bias - was replicated in Study 2, a pre-registered study. Finally, ex-Gaussian analyses of the results of Study 1 and Study 2 helped identify differences between the current research and previous studies that had reported participant sex differences.

Keywords: Facial Expressions, Stereotypes, Evaluative Associations, Social Categorization

When Men are Wonderful: A Larger Happy Face Facilitation Effect for Male (vs Female) Faces for Male Participants

Humans rapidly attribute emotional states to people based on various cues including facial expressions. The ability to recognize facial expressions was initially considered (Bruce & Young, 1986) relatively independent of other face processing abilities such as the recognition of face sex. However, studies have shown that when people are asked to make speeded facial expression decisions, social category information including face sex influences expression recognition (Becker et al., 2007; Bijlstra et al., 2010; Brooks et al., 2018; Craig & Lipp, 2018; Hugenberg & Sczesny, 2006; Smith et al., 2017). For example, in an initial study (Hugenberg & Sczesny, 2006) designed to test for the influence of face sex on speeded expression recognition, participants were asked to repeatedly categorize faces from a standardized set (Ekman & Friesen, 1976) as either happy or angry. The faces were equally often male and female. Even though participants were not instructed to respond to the sex of the faces, face sex exerted an influence on reaction times. Specifically, ANOVA results showed that 1) participants were faster to categorize faces as happy compared to angry and 2) the facilitation effect for happy faces was larger in magnitude for female, compared to male, faces.

According to the authors' (Hugenberg & Sczesny, 2006) interpretation, the influence of face sex on angry and happy recognition is best understood in the context of the Happy Face Advantage (HFA) — the ubiquitous finding that responses are nearly always faster and more accurate responses for happy compared to other expression types. Basing their arguments on separate research (Leppänen et al., 2003; Leppänen & Hietanen, 2003) the authors argued that the HFA is influenced by both the pleasantness of the experimental setting and the pleasantness of happy expression. The HFA is larger for female faces because even though both males and females are both rated positively (Eagly & Mladinic, 1989) females are typically rated more positively – the "women are wonderful" effect — and therefore, the evaluative association between "female" and "pleasant" is stronger leading to faster responses to female-happy faces.

The authors of the evaluative association hypothesis (Hugenberg & Sczesny, 2006) also considered whether stereotypical expectancies for facial expressions, rather than evaluative differences, might be better able to account for RT differences. To do this they compared the effects of face sex on expression recognition for both happy vs angry expressions and happy vs sad expressions. Sadness was chosen because in a previous study (Plant et al., 2000) people rated sadness as more typical of women and therefore, the authors predicted that if an expectancy-based mechanism was responsible for the effects reported in their first experiment then the larger HFA for female faces should disappear when sad facial expressions are selected as the comparison category. In contrast, because sad expressions are negative in valence, the evaluative association account predicts the same pattern for happy vs sad expressions compared to happy vs angry expressions. In Experiment 2, the authors compared all 3 expression types analyzing their results in a 2 (target sex: male; female) × 2 (expression valence: positive; negative) × 2 (negative expression: anger; sadness) mixed-model ANOVA. The 3-way interaction was not significant and consequently, the authors averaged across negative expression types and reported a smaller HFA for male compared to female faces. Overall, the results were interpreted as supporting their hypothesis that evaluative associations and not stereotypes drive the larger HFA for female faces.

Following the initial work of Hugenberg and colleagues, a later study (Bijlstra et al., 2010) showed that gender stereotypes for emotions — expectations about typical emotions expressed (or experienced) by males and females — might influence the expression recognition when differences in valence are made less salient. Participants were assigned to one of 3 conditions 1) dual-valence: 'happy versus angry' 2) dual-valence: 'happy vs sad' and 3) single-valence: angry versus sad. The single valence condition was included as a critical test of their hypotheses because the opposite valence happy face condition has been removed and consequently, valence (positive vs negative) differences were less salient. In their dual valence condition, the authors replicated the absence of differences reported by Hugenberg and Sczesny (Experiment 2; 2006). In the single valence condition, the mean differences followed the pattern predicted for stereotype activation; participants were faster in recognizing anger than sadness on male targets, whereas, for female targets, the opposite pattern was recorded.

What is missing from the analyses conducted so far is an assessment of the magnitude of participant sex differences and more specifically, estimates for male participants. Hugenberg and Sczesny (2006) sampled 47 male participants assigned to 2 between-subjects conditions whereas Bijlstra et al., (Bijlstra et al., 2010) sampled 36 male participants who were randomly assigned to 1 of 3 between subjects conditions described above. Rather than report estimates of effect size estimates for the participant sex separately, the researchers used Null Hypothesis Significance Testing to evaluate the contribution of participant sex differences to speeded expression recognition. In this approach, when participant sex differences fail to reach a threshold (e.g., $p < .05$) the variable participant sex is removed - means are aggregated across participant sex. A problem with focussing on NHST is that, across studies, participant sex differences might make a substantive contribution to the magnitude of an effect (and even determine the sign of differences) but there is no way of assessing this because neither effect sizes nor uncertainty estimates are provided for a single study. A non-significant p-value is not evidence for the absence of an effect - at best it can be considered indeterminate. A better approach is to report effect sizes and associated indices of uncertainty along with threshold statistics (p-values, Bayes Factors etc). Considering the small number of male participants used in past research, the first objective of the current research is to estimate participant sex differences using a larger sample of male participants.

Establishing whether participant sex moderates the effects of face sex on happy and sad expression recognition is critical because the contrast has been used to support an evaluative association over a stereotype account. The negligible effect reported for male participants (Tipples, 2019, 2022a, 2022b) categorising angry and happy expressions indicates that evaluative associations are weaker for male participants. Such effects are not without precedent. For example, reaction time studies (Nosek & Banaji, 2001; Richeson & Ambady, 2001; Rudman & Goodwin, 2004) have also recorded an in-group preference or implicit attitude favouring females. In one study (Rudman & Goodwin, 2004) male participants implicit attitudes were non-significant — a finding interpreted as indicating a neutral implicit attitude (Rudman & Goodwin, 2004). Moreover, in one rating study (Eagly & Mladinic, 1989), the authors found that although both males and females tended to ascribe favourable traits to women (the “women are wonderful”) the effect was larger in magnitude for female participants.

A recent multiverse analyses, indicates that previous studies may have underestimated the influence of participant sex differences due to the type of distribution assumed for the data and, the method used to remove outliers. Multiverse analysis (Steegen et al., 2016) refers to the comparison of different data analyses choices applied to a single dataset and is useful approach for considering the robustness of research findings. The multiverse was applied to RT data from task that required participants to categorise both male and female happy and angry expressions. Nine outlier removal methods were compared across 5 distribution types, 4 of which (the ex-Wald, ex-Gaussian, shifted Wald, and Wiener/Diffusion) were selected for their suitability for RT analyses. The outlier removal methods included the same methods used in past research in this topic (Bijlstra et al., 2018, 2019; Hugenberg & Sczesny, 2006; Tipples, 2019) as well as recommended methods for removing RT outliers (Cousineau & Chartier, 2010; Leys et al., 2013; Voss et al., 2015).

The focus of the multiverse analysis was effect size estimation for the face sex X expression X participant sex interaction term. The multiverse analysis was applied to both aggregated and non-aggregated RTs. For aggregated data, the 9 outlier removal methods were compared for aggregated mean RTs and drift rates (described below). For non-aggregated, trial-level responses, the ex-Wald, ex-Gaussian, shifted Wald and Gaussian Models were estimated as Generalised Linear Mixed Effect Models (GLMMs). GLMMs permit the modelling of by-stimulus random intercepts and slopes as well as the more usual by-participant effects found in ANOVA. The advantage of the latter over the usual ANOVA is that the results are generalizable beyond the sample of stimuli used in the experiment (DeBruine & Barr, 2021; Judd et al., 2012). Now, I will briefly describe the different models.

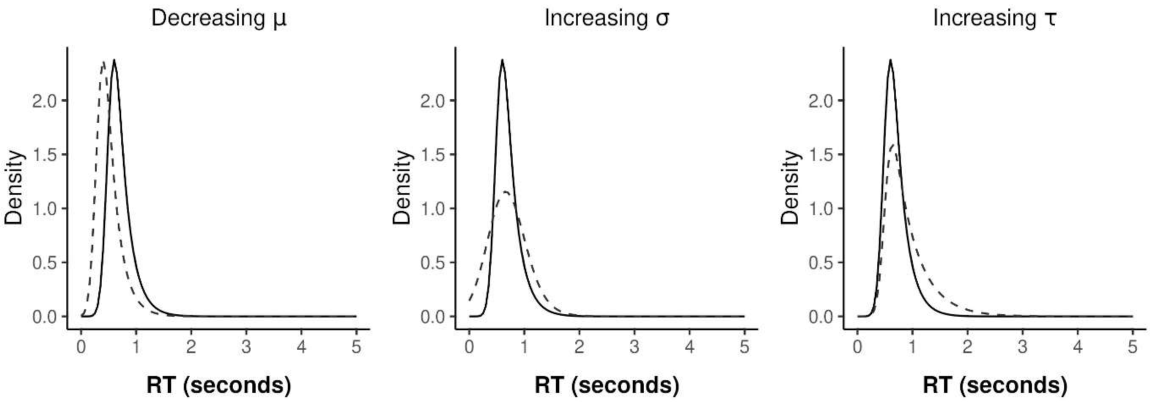

Figure 1. An example of changes in the 3 parameters of the ex-Gaussian. Broken (dashed) lines illustrate a decrease (leftward shift) in mu (left), increased sigma (middle) and increased tau, the exponential component (right).

The ex-Gaussian typically provides an excellent fit to RT data (Luce, 1986) and has helped reveal participant sex differences not observed in mean reaction times (Tipples, 2022b). The ex-Gaussian is illustrated in Figure 1 with changes in parameters visualized as a dashed line. The ex-Gaussian distribution is a convolution of the Gaussian and exponential distributions where the parameters μ (mu) and σ (sigma) are the mean and the standard deviation of the Gaussian component and τ (tau) is the mean and standard deviation of the exponential component. Tau accounts for slow RTs that produce the long tail of the RT distribution. An early shift in the distribution is captured via the mean of the Gaussian parameter mu (Figure 1 – left) whereas late differences will manifest in terms of changes in tau (Figure 1 – right). Changes in sigma (Figure 1 – centre) reflect increased variability in RTs and are expected to follow the pattern for mu with, for example, the more difficult conditions leading to more variable RTs (larger sigma values).

The ex-Wald replaces the Gaussian component of the ex-Gaussian with the Wald (inverse Gaussian) distribution. Therefore, like the ex-Gaussian, the ex-Wald includes 3 parameters with μ and σ, referring to the mean and standard deviation of the Wald portion and τ referring to the exponential portion of the ex-Wald distribution. For RT tasks, both the ex-Wald and shifted-Wald have been proposed as simple accumulation-to-bounds models of cognitive processes. For example, when conceptualized as an evidence accumulation model, the 3 parameters of the shifted-Wald distribution become γ (gamma), α (alpha) and θ (theta) where gamma is the rate at which evidence accumulates toward a single correct response threshold (alpha) and theta describes non-decision processing time before and after the accumulation of evidence. Such models highlight the fact that RT facilitation is not a unitary phenomenon. For example, in the shifted-Wald, RT facilitation might occur due to either increased information accumulation (e.g., due to stronger evaluative associations) or reduced non-decision times (e.g., due the faster encoding of specific facial characteristics) or reduced response caution (e.g., due to a decision by participants to trade accuracy for speed).

The Drift Diffusion Model (DDM; Ratcliff, 1978) also includes parameters for non-decision times, evidence accumulation and a response threshold. A key advantage of the DDM is that the model is applied to all the data – error responses are not excluded from the model. Including error responses means that researchers can estimate differences in the starting point of the evidence accumulation process as well as a more valid estimate of the decision-making threshold namely, the distance between the upper (e.g., "happy") and lower (e.g., "sad") response boundaries. The starting point of the evidence accumulation process might be biased if, for example, the participant receives an incentive to detect happy faces over sad faces. A practical limitation of the DDM is that, when the goal is model more parameters than the drift rate, then the DDM requires a relatively high % of error responses. This makes the model difficult to apply when participants are required to make relatively easy decisions such as deciding whether a face appears either sad or happy.

Summary of Multiverse and ex-Gaussian Model Results

For the analyses of mean RTs using ANOVA, the results of the multiverse (Tipples, 2022a) showed that the application of one recommended outlier removal method based on the Median Absolute Deviation (Leys et al., 2013) effectively removed the skew of the distribution (Skewness coefficient = 0.79) and Bayes Factor analysis indicated "extreme evidence" favouring the inclusion of the critical face sex X expression X face sex interaction. In contrast, when another widely used outlier removal method (e.g., Craig et al., 2018) based on SDs was used, the RT distribution remained skewed (Skewness coefficient = 1.93) and Bayes Factor indicated inconclusive evidence for a model either with or without the term (BF10 = 0.33 to 1). For the analysis of aggregated drift rates from the Diffusion Model, evidence favouring the inclusion of the 3-way interaction was decisive when a method of outlier removal recommended for Diffusion Modelling (e.g., Voss et al., 2015) was applied even though the distribution remained skewed after this method. For the non-aggregated, trial level data, effect sizes were larger and confidence intervals narrower when RT data were modeled using distributions known to provide a good account of RT data namely, the ex-Gaussian, ex-Wald and shifted-Wald distributions.

A further extended ex-Gaussian analysis in which the expression X face sex X participant sex interaction was estimated for all 3 parameters. For the parameter mu, results showed that for female participants there was a 39 ms facilitation effect for female-happy compared to female-angry expressions that reduced in magnitude by an estimated 33 ms for the male-happy vs male-angry contrast. For male participants, the interaction was 30 ms smaller (3 ms in magnitude). Further analyses showed that the female-happy face facilitation for female participants was limited to mu, the location parameter of the ex-Gaussian indicating a fast-acting effect of face sex for female participants. For the parameter tau, for female participants, there was the usual Happy Face Facilitation effect but little evidence of moderation by face sex.

The above ex-Gaussian results were replicated in further research (Tipples, 2022b) in which the author applied the ex-Gaussian model to published RT data gathered by a separate research group (Craig et al., 2018). The authors of the research had designed their study to investigate the influence of face sex and face race and on the speeded recognition of angry and happy faces. Ex-Gaussian modelling revealed interactions between face race, face sex, expression type and participant sex that were not significant for when data were analysed by the original authors using either a linear mixed effect model or analysis of mean RT using ANOVA. Specifically, the ex-Gaussian model results showed that for female participants, the size of the face sex X expression type interaction effect (for the faces of white individuals) was nearly identical to the research described above (Tipples, 2022a) with a 39 ms facilitation effect for female-happy (vs female-angry) faces that was reduced to 4 ms for male faces. A different pattern for female participants responding to the faces of black individuals. For female participants, the mean of the Gaussian component (mu) revealed faster RTs to the faces of black, female-angry compared to black, male-angry individuals. However, analyses of tau (the slow component) revealed a cross over pattern with a happy face facilitation effect for the faces of black female individuals and the reverse pattern for the faces of black male individuals. A facilitatory-inhibitory process account was proposed to explain female participants to the faces of black individuals (for further details see; Tipples, 2022b).

Interim summary

The multiverse and ex-Gaussian model results from studies comparing angry and happy expressions suggest that for female participants show an ingroup bias when responding to the faces of white individuals. This specific pattern may reflect the stronger affiliative needs of female participants (Taylor et al., 2000) who respond to male displays of anger by seeking warmth and connection to an ingroup member — this affiliative need increases the evaluative associations between “female” and “positive” leading to facilitation effect for female-happy expressions in female participants.

The multiverse and ex-Gaussian model results also highlight the value of considering the distribution of RT data when testing for participant sex differences. For the critical comparison between sad and happy expressions — used to support the evaluative association over the stereotype account — the authors of one study (Bijlstra et al., 2010) did address the positive skew of the RT distribution when they applied a log-transformation. However, as already noted, critical issues remain namely are the small number of male participants. Furthermore, log-transformation does not allow the separation of slow from fast responses and consequently, slow and fast responses are analysed together in a single model.

Current Research

Considering the ex-Gaussian analyses (Tipples, 2019; Tipples, 2022a; Tipples, 2022b) multiverse findings and other research reporting ingroup bias specific to female participants (Eagly & Mladinic, 1989; Nosek & Banaji, 2001; Richeson & Ambady, 2001; Rudman & Goodwin, 2004), I wanted to establish whether participant sex might also moderate a key finding reported previously namely, the influence of face sex on sad and happy facial expression recognition. The small number of male participants and the decision to aggregate across participant sex may mean that researchers have overlooked the contribution of sex differences. The sad vs happy comparison matters because as noted above, the combination of an absent

RT facilitation effect for female-sad vs male-sad expressions and a larger Happy Face Advantage for female faces was used to support the evaluation association hypothesis over the stereotype account. In other words, this effect is critical for the argument that stronger positive evaluative associations for female faces are responsible for the effect of face sex on expression recognition.

With respect to the choice of model for RTs, I took several approaches. First, I analysed both mean proportion of correct responses and mean correct log-transformed RTs using traditional ANOVA. For the RTs, I followed past research (Bijlstra et al., 2010) using the RT cut-offs "> 200 & <3000 ms" and applying a log-transformation. As noted above, the log-transformation can be an effective approach to addressing the skew of the RT distribution even though it is not always applied for the analysis of expression decision times (e.g., Hugenberg & Sczesny, 2006). Participant sex differences are predicted for ANOVA of log-transformed RTs with a larger happy face facilitation effect for female faces vs male faces for female participants and a reduction in this effect for male participants.

Second, following recent research, I conducted a multiverse analysis of the same 9 outlier methods used previously. These outlier methods were applied to both mean aggregated RTs and in addition, mean log-transformed RTs. Again, the goal was to establish whether evidence supported the inclusion of the face sex X expression X participant sex interaction across different outlier methods. The prediction for multiverse is, following recent research (Tipples, 2022a), that the face sex X expression X participant interaction will be largest for the outlier removal approach and RT transformation that effectively removes the skew of the RT distribution. Tipples (2022a) did not include log-transformed RTs in the multiverse analysis. Log-transformation might be expected to correct the skew of the distribution and therefore, one possibility is that analyses support the inclusion of the interaction term for the multiverse of log-transformed data irrespective of the outlier approach used.

Third, to facilitate comparison with recent ex-Gaussian modelling results I have included, in separate section, an extended ex-Gaussian analysis. Following previous research (Tipples, 2019, Tipples, 2022a; Tipples, 2022b), the prediction for ex-Gaussian analysis is that there will be a larger happy face facilitation effect for female (vs male) faces in female participants for the location parameter mu. The interaction term is predicted to indicate a significant reduction in this effect for male participants. Put differently, the larger effect for female participants will manifest among the fastest RTs captured by mu. In addition to permitting analysis of the time course of effects, a specific advantage of the GLMM variant of the ex-Gaussian applied here (compared to ANOVA) is that by-item variability can be included in the model permitting generalisation beyond the selection of faces of individuals used in the study.

Study 2 is a pre-registered attempt to replicate the unexpected pattern reported for male participants in Study 1. For Study 2, I had pre-registered regression modelling the RT data using the same shifted-Wald analysis used recently (Tipples, 2022a). I conducted the ANOVA of log-transformed RTs in response to a reviewer request and therefore, for the sake of parsimony, the shifted-Wald results are reported in Appendix A. The shifted-Wald results are identical to the ANOVA of log-transformed RT data.

Study 1

Method

Overview. To test the key hypothesis, I used the same stimulus sets, the same task, and the same number of male participants that I used in recent research (Tipples, 2022a) designed to study participant sex differences. The main difference was the presentation of sad facial expressions rather than angry facial expressions and the requirement for participants to press the S key (for sad expressions) rather than the A key (for angry expressions). All results were gathered between 2021 and 2022.

Sample size

Sample size estimation was based on the ex-Gaussian analysis reported in past research (Tipples, 2022a). In the latter research, the t-value for the 3-way interaction term was – 4.12. This represents a large between-subject effect (Cohen's d_z = -0.91) and consequently, power analyses (Faul et al., 2007) indicates that as few as 21 participants for each participant sex group are required to achieve power .80 with alpha set to .05. Nonetheless, I continued sampling until a sample size of 34 male participants was reached. This provides sufficient power (beta = .80; alpha = 0.05; one-tailed) to estimate a medium-sized expression (happy) X face sex (male) slope (within-subjects t-test) for male participants.

Participants. Eighty-six students from Leeds Beckett University took part in the study in return for course credit. The final sample consisted of 34 males (Age: *M* = 26, *SD* = 11) and 52 females (Age: *M* = 23, *SD* = 12). Before commencing the study, ethical approval was obtained from the ethics committee of the University Ethics Committee.

Face Stimuli. The faces were of 34 individuals (17 females, 17 male) each displaying happiness and sadness (68 faces in total) that were selected from 3 face databases: 1) Karolinska Directed Emotional Faces (KDEF; Lundqvist et al., 1998) 2) the Pictures of Facial Affect (POFA; Ekman & Friesen, 1976) and 3) NimStim (NIM; Tottenham et al., 2009). The NIM and KDEF sets each consisted of the faces of 6 male and 6 female individuals each displaying one happy and a sad expression. The POFA set consisted of the faces of 5 male and 5 female individuals each displaying one happy and a sad expression. The face images were scaled (in proportion) to 424 pixels in height.

Procedure. Participants completed 4 blocks of 68 trials separated by a brief rest period after trial number 136. Each block was composed of equiprobable factorial combinations of face sex (male, female) and facial expression (sad, happy). A new randomized trial order sequence was created for each block, for each participant, based on a computer-generated random seed. Sixteen practice trials using faces not used in the main experiment preceded the first main block of trials.

The trial sequence was: 1) 1000 milliseconds blank interval 2) 500 milliseconds fixation cross and 3) the face stimulus until either a participant made a response, or 3.5 seconds had elapsed. If participants failed to respond within 3.5 seconds, they received the feedback “too slow” for an extra 500 milliseconds. Participants were instructed to respond as quickly and accurately as possible. Participants responded by pressing the S key with their left index finger to indicate a sad expression and the H with their right index finger to indicate a happy expression. This study was not pre-registered. Data and code can be found here https://osf.io/p35v7/ .

Results

Pre-treatment and outlier removal. The criterion for including a participant's dataset was 1) an overall % accuracy rate greater than 60% and 2) an overall mean RT less than 1.2 seconds. A

ANOVA of RT and Accuracy Rates

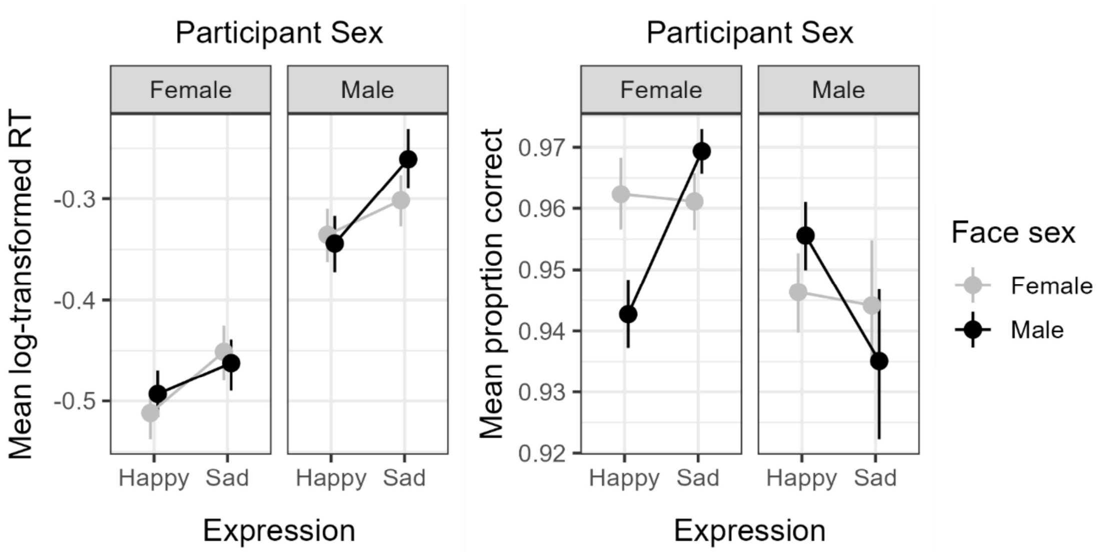

Figure 2. The mean log-transformed RTs (left) and proportion correct (right) as a function of expression (happy, sad) and face sex (female, male) for female and male participants separately. Error bars are boot-strapped standard errors.

RTs. Following previous research (Bijlstra et al., 2010) incorrect trials (4.3%) and response latencies below 200 ms or above 3000 ms (4.4%) were excluded and the RTs were subsequently log-transformed. The mean correct log-transformed RTs were analyzed in a 2 X 2 X 2 mixed ANOVA with expression (happy, sad) and face sex (male, female) as within subject variables and participant sex (male, female) as the between-subjects variable. The pattern of estimated means displayed in Figure 2 (left) suggests that the effects of face sex on sad and happy facial expression recognition differ between male and female participants. A 3-way face sex X expression X participant sex interaction effect, $F(1, 73) = 17.98$, $p < .001$ ($\eta^2_p = 0.20$) supports this observation. ANOVA also revealed a main effects of participant sex, $F(1, 73) =$

21.83, $p < .001$ (η^2_p = 0.23) with overall faster RTs for female participants compared to male participants , a main effect of expression, $F(1, 73) = 43.20$, $p < .001$ ($\eta^2_p = 0.37$) with faster RTs to happy compared to sad expressions and a statistically small ($\eta^2_p = 0.06$) main effect of face sex, $F(1, 73) = 4.77$, $p = .03$. Simple interaction effect analyses of the 3-way interaction revealed significant face sex X expression interaction effects for both male participants $F(1, 33) = 12.68$, $p = .001$ (η^2p = 0.28) and female participants $F(1, 40) = 5.69$, $p = .02$ ($\eta^2_p = 0.12$).

Simple Main Effects. In terms of the happy vs sad difference — used previously to test for moderation of the happy face advantage — results showed that for female participants were faster to judge female faces as happy compared to sad, Cohen's $d_z = 0.84$, 95% CI [0.49; 1.22] and this happy-sad difference was reduced in magnitude for male faces, Cohen's $d_z = 0.40$, 95% CI [0.09; 0.73]. For male participants, the happy face facilitation effect was larger for male faces, Cohen's $d_z = 0.83$, 95% CI [0.46; 1.26] compared to female faces, Cohen's $d_z = 0.44$, 95% CI [0.10; 0.81]. Focusing on face sex differences for male participants, results indicated faster RTs to male-sad compared to female-sad expressions, Cohen's $d_z = 0.58$, 95% CI [0.23; 0.97] whereas for female participants, the same difference was of the opposite sign, and not significant Cohen's $d_z = -0.15$, 95% CI [-0.47; 0.15].

Accuracy. Following the RT data there was a 3-way face sex X expression X participant sex interaction, $F(1, 73) = 13.33$, $p < .001$ ($\eta^2_p = .15$) and a 2-way expression X participant sex interaction $F(1, 73) = 4.80$, $p = .03$ ($\eta^2_p = .06$). All other effects were not significant (largest η^2_p =. 04). Simple interaction effect analyses of the face sex X expression X participant sex interaction showed that the face sex X expression was and significant for female participants, $F(1, 40) = 9.97$, $p = .003$ ($\eta^2_p = .20$) and male participants, $F(1, 33) = 4.22$, $p = .04$ ($\eta^2_p = .11$). For female participants, accuracy was lower when responding to male-happy ($M = 0.94$; $SD = 0.03$) compared to male-sad ($M = 0.96$; $SD = 0.02$) expressions, Cohen's $d_z = 0.66$; 95% CI [0.33; 1.02]. For females, difference in accuracy between the female-happy ($M = 0.96$; $SD = 0.03$) and female-sad ($M = 0.96$; $SD = 0.03$) faces was small and not significant Cohen's $d_z = -0.02$; 95% CI [-0.33; 0.28]. For male participants, accuracy was lowest for male-sad expressions ($M = 0.93$; $SD = 0.07$) compared to the remaining face types although, as shown in Figure 2 (right) all contrasts were small in magnitude and not significant including the male-happy vs male-sad difference, Cohen's d_z = -0.28; 95% CI [-0.64; 0.05].

Multiverse Analysis

	BF Inclusion	p	Skewness
Outlier Removal		Mean RT (a)	
1) <2.5 SDs	8.46	.0006	2.75
2) >200 ms & < 3000 ms	82.62	.0002	2.54
3) > 100 ms & < 3 SDs	54.26	.0008	2.79
4) >200 ms & < 2500 ms	160.15	.0001	2.22
5) 3*MAD	18.41	.0002	0.78
6) >1 ms	13.47	.0019	3.42
7) 3*IQR(log(RT))	30.96	.0010	3.18
8) transformed	8.86	.0004	1.01
9) >100 ms	15.19	.0019	3.42
		Log RT (b)	
1) <2.5 SDs	24.02	.0002	0.75
2) >200 ms & < 3000 ms	13.59	.0001	0.94
3) > 100 ms & < 3 SDs	138.26	.0002	0.88
4) >200 ms & < 2500 ms	111.29	.0001	0.86
5) 3*MAD	14.17	.0003	0.17
6) >1 ms	35.83	.0002	1.02
7) 3*IQR(log(RT))	28.9	.0002	1.03
8) transformed	10.36	.0003	0.35
9) >100 ms	35.93	.0002	1.11

Table 1. Bayes Factors (BF10) inclusion results and p-values for face sex X expression X participant sex interaction term for different outlier removal criterion for mean RTs (top – a) and log-transformed RTs (bottom – b). The final column is the skewness statistics for the different outlier removal criterion.

Outlier Removal	Removed %
1) <2.5 SDs	3.43
2) >200 ms & < 3000 ms	1.62
3) > 100 ms & < 3 SDs	3.16
4) >200 ms & < 2500 ms	1.8
5) 3*MAD	7.77
6) >1 ms	1.29
7) 3*IQR(log(RT))	1.65
8) transformed	5.57
9) >100 ms	1.31

Table 2. The % of responses removed for each of the 9 outlier methods

Table 1 shows the results of the Multiverse analyses. The % of responses removed for each outlier method are presented in Table 2. BF Inclusion (Mathot, 2017) refers to the Bayes Factor for including the 3-way interaction term relative to models without the interaction term. Following Jeffreys (Jeffreys, 1961) Bayes Factors ranging from 3 to 10 represent substantial evidence, 10 to 30 strong evidence, 30 to 100 very strong evidence and finally, "> 100" represents decisive evidence for including the 3-way interaction term. In contrast to the multiverse reported in recent research (Tipples, 2022a), Bayes Factor analyses supported the inclusion of the 3-way interaction term irrespective of the type of outlier method applied for both Mean RTs (Table 1 – top) and log-transformed RTs (Table 1 – bottom). Moreover, and in contrast to previous research (Tipples, 2022a), focusing on mean RTs (Table 1 – top) the Bayes Factor inclusion value for the outlier method that in the lowest skew coefficient (.78) was lower (*BF* = 18.41) compared to other methods that left the distribution relatively skewed (e.g., >200 ms & < 2500 ms; *BF* = 160.15; Skewness = 2.22). Another way to view the multiverse results is to note that evidence favoring the inclusion of the 3-way interaction was weaker when more of the distribution was removed. For example, the 2 currently recommended approaches for outlier removal removed 7.7% and 5.5% of the total responses resulted in the lowest BF inclusion values.

Discussion

The results are further support for participant sex differences in the effect of face sex on expression decision times. For female participants, there was a larger happy face facilitation effect for female vs male faces. However, rather than an expected reduction of the size of this effect for male participants, results indicate a reversal of the pattern reported for female participants — for male participants, the happy face facilitation effect was larger in magnitude for male compared to female faces. In short, to the extent that such differences reflect evaluative associations then the pattern reflects an overall ingroup bias for both participant sexes

Study 2

The results for male participants in Study 1 contradict the often-replicated pattern of a larger happy face facilitation effect for female compared to male faces. Considering the novelty of the findings, I wanted to replicate the pattern for male participants in a new sample of male participants. Therefore, I pre-registered a replication attempt of Study 1. One deviation from the pre-registration is that in Study 2 I have reported the same ANOVA of mean log-transformed RTs reported for Study 1 with the original pre-registered analysis (a shifted-Wald model) reported in Appendix 1. The shifted-Wald model results are identical to that reported for the ANOVA of mean log-transformed RTs. Based on the results of Study 1, the key prediction for male participants is the same pattern reported for Experiment 1 namely a larger happy-face-facilitation effect for male compared to female faces.

Method

Power Analyses. Thirty-four male participants were recruited using the online platform, Prolific Academic (https://prolific.ac). A sample size of 34 provides sufficient power (beta = .80; alpha = 0.05; one-tailed) to estimate a within-subjects t-test for male participants.

Participants. The sample consisted of 34 males (Age: *M* = 26, *SD* = 7.8). The sample consisted of 26 white, 4 black and 1 male of unidentified ethnic origin. The 4 black males reported a South African national identity, and the remaining (white) participants reported the following nationality identity from Estonia, Greece, Hungary, Israel, Italy, Mexico, Poland, Portugal and the United Kingdom. Before commencing the study, ethical approval was obtained from the ethics committee of the University Ethics Committee.

Face Stimuli. The faces were identical to those used in Study 1.

Procedure. Participants completed 5 blocks of 68 trials with brief rest period after the trial 136 – an additional block of 68 trials was added due to a programming error leading to the creation of 340 trials rather than the 272 used in Study 1. Again, each block was composed of equiprobable factorial combinations of face sex (male, female) and facial expression (sad, happy). A new randomized trial order sequence was created for each block, for each participant, based on a computer-generated random seed. Sixteen practice trials preceded the first main block of trials. The trial sequence was identical to that reported in Study 1. Data, Code and pre-registration can be found here https://osf.io/p35v7/ .

Results

Following Study 1, the criterion for including a participant's dataset was 1) an overall % accuracy rate greater than 60% and 2) an overall mean RT less than 1.2 seconds. The dataset of 1 participant was removed (mean R*T* = 1.36) leaving 33 male participants.

Traditional ANOVA of RT and Accuracy Rates

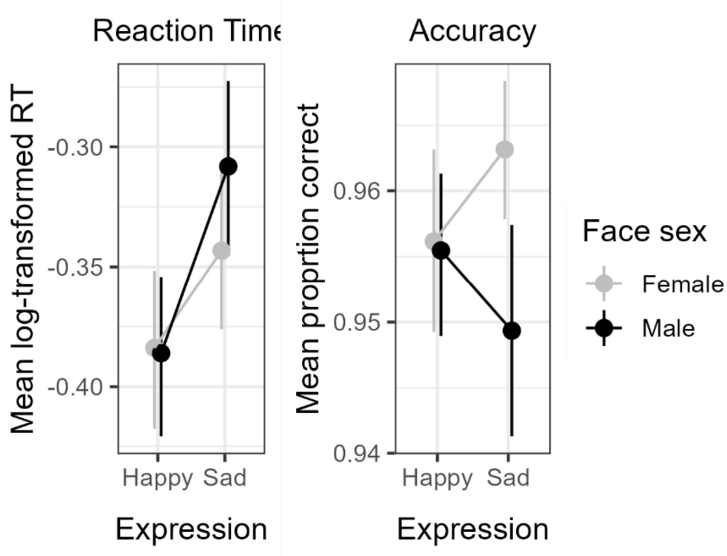

Figure 3. The mean log-transformed RTs (left) and proportion correct (right) as a function of expression (happy, sad) and face sex (female, male) for male participants in Study 2. Error bars are boot-strapped standard errors.

RTs. Following Study 1 and previous research (Bijlstra et al., 2010) incorrect trials (4.0%) and response latencies below 200 ms or above 3000 ms (4.1%) were excluded and the RTs were subsequently log-transformed. The mean correct log-transformed RTs were analyzed in a 2 X 2 repeated measures ANOVA with expression (happy, sad) and face sex (male, female) as within subject variables. The pattern of estimated means displayed in Figure 3 (left) and replicates the pattern for males reported in Study 1 namely, a 2-way interaction between face sex and expression, $F(1, 73) = 21.83$, $p < .001$ ($\eta^2_p = 0.23$) with the HFA for female faces, Cohen's d_z = 0.49, 95% CI [0.15; 0.87] increasing in magnitude for male faces, Cohen's d_z = 0.90, 95% CI [0.52; 1.34].

Accuracy. The mean proportion of correct responses were analyzed in a 2 X 2 repeated measures ANOVA with expression (happy, sad) and face sex (male, female) as within subject variables. All effects were not significant, (expression, $F(1, 33) = 0.00$, $p = .95$ ($\eta^2_p < 0.01$); face sex, $F(1, 33) = 2.83$, $p = .10$ ($\eta^2_p = 0.08$); expression X face sex, $F(1, 33) = 1.63$, $p = .23$ ($\eta^2_p = 0.05$))

Discussion Study 1 and Study 2

The results of Study 1 and Study 2 converge on the same conclusion namely, the previously reported larger facilitation effect for female compared to male faces is restricted to female participants. The novel result is to show that for male participants the happy face facilitation effect is larger male compared to female faces. The is effect was reported for both Study 1 and a second pre-registered study (Study 2). As already noted, this is not the pattern reported in past research that reported moderation of face sex X expression by participant sex. Moreover, the results of the multiverse also differ that reported recently (Tipples, 2022a) with all Bayes Factors supporting the inclusion of the 3-way interaction term, irrespective of outlier method.

Study 3: Ex-Gaussian Modelling of the Results of Study 1 and Study 2

To facilitate comparison between the research and the results of recent research I also conducted ex-Gaussian analyses. The novel information provided by the ex-Gaussian analyses is that provides an analysis of slow and fast components. Therefore, this analysis has the potential to clarify why the results of multiverse analysis reported here differ from those reported in recent research (Tipples, 2022a). The multiverse analysis reported for Study 1 indicates that removing more of the distribution weakens the 3-way interaction. This could be due to larger effect in the tail of the distribution for either male or female participants. Applying the ex-Gaussian will also permit a replication of the time course effect for female participants reported in previous research. In previous research, for female participants, the reduction in the happy face facilitation effect for male faces was restricted to mu. Mu is the location parameter of the ex-Gaussian and describes the fastest responses from the RT distribution.

Study 1

The ex-Gaussian analyses was conducted using the GAMLSS package (Rigby & Stasinopoulos, 2005) written in the language R (R Core Team, 2020). Following recent research (Tipples, 2022a, Tipples 2022b), incorrect responses and RTs that were either less than 100 ms or more than 3500 ms were removed from the dataset. The regression modelling used in GAMLSS permits the direct estimation of effects of interest using interaction contrast codes (for a tutorial see; Schad et al., 2020). The key prediction namely, that the happy face facilitation will be larger for female faces in female participants for the parameter mu, can be estimated in a single interaction contrast using the default treatment (or dummy) contrast coding in R. Specifically, the b expression (happy=1, sadness=0) X face sex (male=1, female=0) X participant sex (male=1, female=0) was regressed onto mu, sigma, and tau in a single, multilevel regression model. For the parameter mu, the intercept for this model is the estimate of RTs for female-sad expressions for female participants. The expression (happy) slope is an estimate of the happy face facilitation effect (happy minus sad) for female participants responding to female faces. The *b* expression (happy=1, sadness=0) X face sex (male=1, female=0) interaction term is an estimate of the change in the happy face facilitation effect for male compared to female faces. For mu, the latter term is precited to be positive indicating a reduction (attenuation) in the happy face facilitation effect for male compared to female faces (for female participants). Finally, the *b* expression (happy=1, sadness=0) X face sex (male=1, female=0) X participant sex (male=1, female=0) is an estimate of the change in the latter interaction effect between male and female participants – the 3-way interaction is predicted to negative indicating a reduction, for male participants, in the magnitude of the happy face facilitation effect for female compared to male faces. In short, in contrast to the F-test reported for ANOVA, all regression coefficients in this model have substantive theoretical meaning.

To calculate model estimates and confidence intervals for male participants, I re-estimated the identical model but with an estimate for female-sad expressions for male participants as the intercept. For the random effects, I attempted to fit the most complex model possible. For by-participants random effects, all models and parameters included varying by-subject intercepts and varying by-subject slopes for face sex and expression and, the face sex X expression interaction. For items (individual face identities) I included a random intercept for face identity. Models with more complex by-item random effects (e.g., a random, by-items slope for the variable "expression") failed to converge.

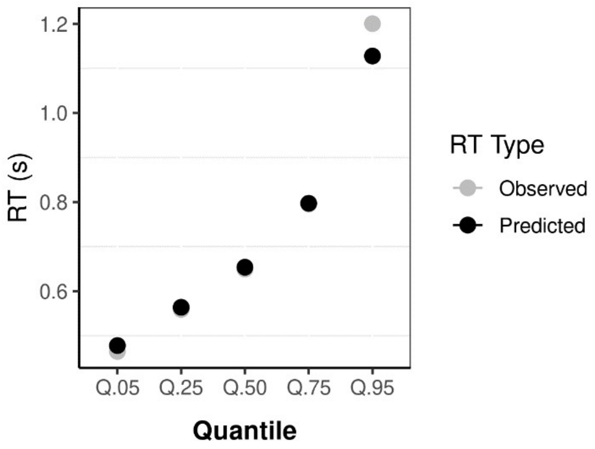

Figure 4. The quantiles for the observed data (grey) plotted against simulated values (black). To create the plot, simulated RTs (10,000 iterations per participant and quantile) were generated from the fitted parameter estimates and then aggregated across participants.

Assessment of Model Quality. A graphical check of model quality is provided in Figure 4 where I have plotted RT quantiles for observed data (grey) plotted against simulated values (black). To create the plot, simulated RTs (10,000 iterations per participant and quantile) were generated from the fitted parameter estimates and then aggregated across participants.

Model Results

Figure 5. The estimated means for female and male participants for the ex-Gaussian parameters mu (left), sigma (middle) and tau (right) as a function of expression and face sex.

Mu. As shown in Figure 5 (left) the pattern of estimated means suggests that for the parameter *mu*, the effects of face sex on speeded sad and happy categorization are qualitatively different for males compared to female participants. This observation is supported by a 19 ms, 3-way interaction effect, b = -0.019, t = -2.77, p = .006, 95% CI[-0.032, -0.005] with significant f. sex (male) X expression (happy) interaction for female participants and a non-significant pattern for male participants. For female participants, results replicate previous research: Female participants were faster to categorize female-happy compared to female-sad expressions, b (happy) = -0.018, t = -6.84, p < .0001, 95% CI[-0.023, -0.012] and this happy-sad difference was reduced by 12 ms for male faces as indicated by the 2-way b (happy) X (f. male) interaction = 0.012, t = 3.24, p = .001, 95% CI[0.005, 0.019]. Female participants were also 7 ms faster to respond to male-sad compared to female-sad expressions, b (f. male) = -0.007, t = -2.75, p = .006, 95% CI[-0.012, -0.002]. For males, the happy face facilitation for female faces that was 5 ms in magnitude and not significant, b (happy) = -0.005, t = -1.15, p = .25, 95% CI[-0.012, 0.003] with a small (non-significant), 6 ms increase in the magnitude of this difference for male faces, b (happy) X (f. male) = -0.006, t = -1.09, p = .275, 95% CI[-0.017, 0.005].

Sigma. For sigma, the b (happy) X (f. male) X (p. male) interaction term was significant = -0.35, t = -2.46, p = .014, 95% CI[-0.63,-0.07]. For female participants, all effects for sigma were not significant. Re-estimating the model with responses made by male participants to female-sad expressions as the model intercept, showed that variability was higher for female-happy compared to female-sad expressions, *b* (happy) = 0.19, *t* = 2.43, *p* = .015, 95% CI[0.037, 0.343] with the interaction term indicating an elimination of this effect for male faces, b (happy) X (f. male) = -0.251, *t* = -2.318, *p* = .02, 95% CI[-0.463, -0.039]. In short, results are broadly consistent with results for mu for male participants – slower RTs to female-happy faces for male participants were associated with higher variability. Put differently, the more "difficult", harder to process conditions were more variable.

Tau. For tau, the 3-way b (happy) X (f. male) X (p. male) interaction term was significant *b* = -0.186, *t* = -3.18, *p* = .001, 95% CI[-0.3, -0.07]. For female participants the data are best represented by a main effect of expression type. As shown in Figure 5 (right), for female participants, tau was smaller for female-happy compared female-sad expressions, *b* (happy) = -0.16, *t* = -5.76, *p* < .001, 95% CI[-0.214, -0.106] with only a small (non-significant) reduction in the magnitude of the effect for male faces, *b* (happy) X (f. male) = 0.05, *t* = 1.37, *p* = .169, 95% CI[-0.023, 0.132]. Re-estimating the model with responses made by male participants to female-sad expressions as the model intercept, showed that happy face facilitation effect for female faces, *b* (happy) = -0.143, *t* = -4.831, *p* < .0001, 95% CI[-0.201, -0.085] increased for male faces *b* (happy) X (f. male) = -0.132, *t* = -3.08, *p* = .002, 95% CI[-0.216, -0.048]. Focusing on face sex differences for male participants, results indicated increases in tau for male-sad compared to female-sad expressions, *b* (f. male) = 0.061, *t* = 2.09, *p* = .036, 95% CI[0.004, 0.119].

Study 2

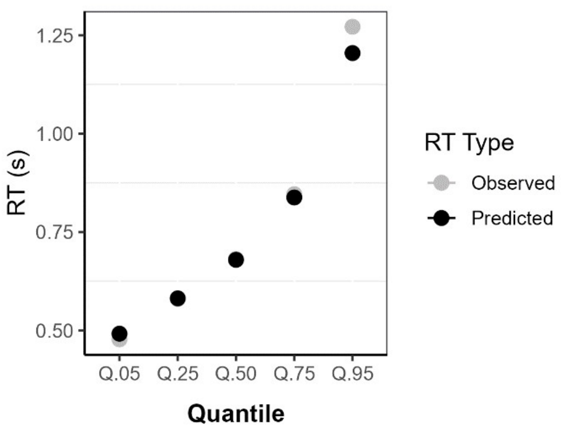

Figure 6. The quantiles for the observed data (grey) plotted against simulated values (black). To create the plot, simulated RTs (10,000 iterations per participant and quantile) were generated from the fitted parameter estimates and then aggregated across participants.

Results

The ex-Gaussian model for Study 2 was identical to that used for Study 1, except that predictor gender was dropped from the regression equation. The 2-way *b* expression (happy) X face sex (male) interaction term created from the treatment coded predictors, expression (0=happy, 1=sad) and face sex (0=female, 1=male) and participant sex (0=female, 1=male) was regressed onto mu, sigma and tau in a single ex-Gaussian regression model. A graphical check of model quality is provided in Figure 6 where I have plotted RT quantiles for observed data (grey) plotted against simulated values (black). To create the plot, simulated RTs (10,000 iterations per participant and quantile) were generated from the fitted parameter estimates and then aggregated across participants. The skewness was low (.02) for the fitted model. Notwithstanding the misfit for the very slowest RTs, the model provided a good fit to the observed data.

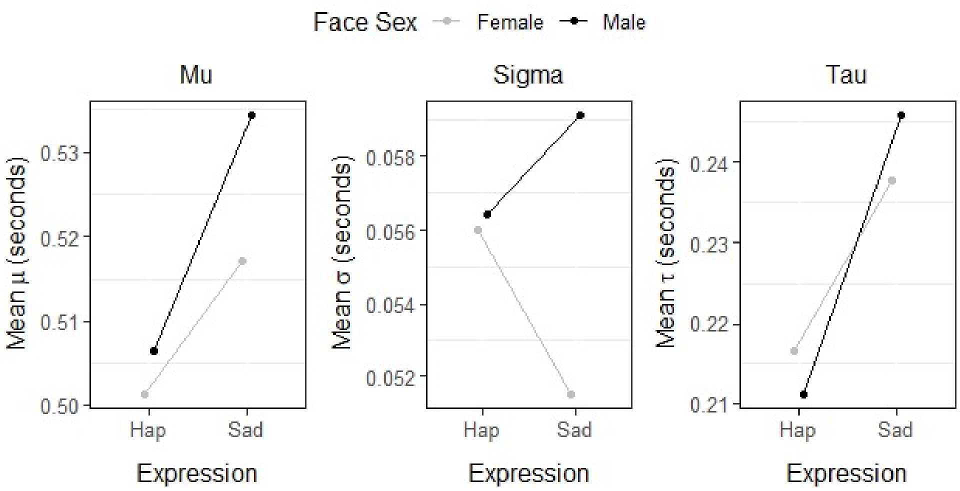

Figure 7. The estimated means for male participants in Study 2 for the ex-Gaussian parameters mu (left), sigma (middle) and tau (right) as a function of expression and face sex.

Mu. For mu, as shown in Figure 7, male participants were an estimated 7 ms faster to categorize female-happy compared to female-sad expressions, *b* (happy) = -0.007, *t* = -2.23, *p* = .02, 95% CI[-0.01, -0.001] and this happy-sad difference increased by 13 ms for male faces as indicated by the 2-way *b* (happy) X (f. male) interaction = -0.013, *t* = -2.69, *p* = .003, 95% CI[-0.02, -0.005].

Sigma. Modelling results for sigma revealed higher variability in RTs for female-happy faces *b* (happy) = 0.14, *t* = 2.31, *p* = .02, 95% CI[0.023, 0.273] that was non-significantly reduced when the faces were male *b* (happy) X (f. male) interaction = -0.15, *t* = -1.67, *p* = .09, 95% CI[-0.33, 0.026].

Tau. For tau, the 2-way b (sad) X (f. male) X (p. male) interaction term was not significant although the pattern is the same as that reported for Study 1 namely, a cross-over pattern (see Figure 7 – right) with faster responses to female-happy expressions compared to female-sad expressions b (happy) = -0.157, *t* = -5.63, *p* < .0001, 95% CI[-.21, -.10] that increases in magnitude for male faces b = -0.045, *t* = -1.13, *p* = .25, 95% CI[-0.12, 0.03].

Discussion

The ex-Gaussian model results help clarify the differences between the current research and recent research (Tipples, 2019, 2022a, 2022b) that applied the ex-Gaussian model to facial expression decision times for happy and angry expressions. Following recent research, the large happy face facilitation effect for female faces for female participants was again restricted to mu, supporting the operation of a relatively fast process. For male participants however, RTs were overall much slower and moreover, the larger happy face facilitation for male faces was found across the RT distribution — it was not restricted to mu, the location parameter that indexes the fastest responses. A further finding is that for female participants the female-happy (vs female-sad) facilitation effect was much smaller than that reported in recent research. I discuss the latter issue in more depth in the General Discussion.

General Discussion

Results are further evidence of participant sex differences in the effects of face sex on speeded expression recognition. For male participants, the happy face facilitation effect was larger in magnitude for male faces compared to female faces and, responses were also faster for female-sad compared to male-sad expressions. This pattern was replicated (Study 2) in pre-registered study of 34 additional male participants. The effect occurred for ANOVA of log-transformed mean RTs using the outlier removal method used in past research (Bijlstra et al.,

2010) and, as shown in the Multiverse Analysis across multiple other outlier removal methods. Finally, the differential pattern for male and female participants was also found for ex-Gaussian analyses.

The results have important implications because the happy vs sad comparison has been considered critical for comparing an evaluative association vs stereotype account. Under the dual-valence conditions presented here, the Happy Face Advantage is predicted to be larger for female compared to male faces. To re-iterate, the idea is that presenting participants with expressions of opposite valence (e.g., sad and happy expressions) - dual valence - leads to an evaluative mindset and consequently, the female-happy bias will dominate effects because females are rated more positively than males. If we accept the evaluative association account, then the interpretation for the results here is that male participants have stronger positive associations for male faces and female participants conversely have stronger positive associations for female faces. In other words, the results suggest ingroup favouritism for male faces in male participants and for female faces in female participants.

One challenge to a pure evaluative association account is that an ingroup bias for male participants was not found in ether recent research conducted by the first author (Tipples, 2019; Tipples, 2022a) or analysis of data collected by other researchers (Tipples, 2022b; Craig et al, 2018). Instead, results of recent research supported a reduction in the magnitude of the same effect for female participants – the effect size was virtually nil in male participants, but it was not reversed. Differences in methodology between the current and past research are unlikely to lie behind the failure to observe an ingroup bias for male participants because the current study used the same large stimulus set used in past research (Tipples, 2022a) with only a minor variation in method (participants pressed "S" for sad expressions rather than "A" for angry). I think that an unmeasured participant characteristic may be the reason for the male-participant ingroup bias reported here. Future studies may benefit from including direct measures of participant gender attitudes including a measure of sexist attitudes.

Effect sizes

For the ex-Gaussian analysis, effect sizes were notably smaller for female participants compared to recent research in which participants categorised angry and happy expressions.

Specifically, in a recent study in which I also applied the ex-Gaussian, the reduction the magnitude of the happy-angry difference for male faces (for mu) was an estimated 33 ms with 95% CIs indicating a narrow range of values compatible with the data (95% CI[0.023, 0.043]). Here, the reduction was 12 ms with 95% CIs also indicating a narrow range of values compatible with the data 95% CI[0.005, 0.019]). Although effect sizes estimates were not provided, Bijstra and colleagues also reported (in a footnote) a marginally significant reduction for male vs female face HFA for happy vs sadness compared to the same difference for happy vs anger.

The smaller effect size for female participants should not be ignored as it may provide important information about the effect. The larger effect of face sex for female participants judging angry and happy expressions is consistent with the idea that for female participants, there exists weaker associations exist between maleness and sadness compared to maleness and anger. Again, male displays of anger are likely to be particularly salient to young female individuals (who form most individuals who have taken part in research on this topic) because, as research shows, male aggression is a particular threat to young versus older females (Shackelford et al., 2003).

In the context of male anger, female faces displaying happiness may meet affiliative needs for young female participants — they are “befriending” signal (Taylor et al., 2000). Consequently, as noted in the introduction, the evaluative associations between “female” and “positive” leading to facilitation effect for female-happy expressions in female participants.

Difference in Methods (Constraints on Generality)

There are several differences between the method used here and in past research (Bijlstra et al., 2010; Hugenberg & Sczesny, 2006) that tested for an effect of face sex on speeded happy and sad facial expression recognition. Most notably, in past research, faces were displayed for 200 ms whereas in the current study were displayed until either the participant made a response, or 3.5 seconds had elapsed. Furthermore, in the current study I also added the feedback “too slow” if the participant failed to respond within 3.5 seconds. The difference in exposure duration between the present and past research (Bijlstra et al. 2010) is the likely reason for the differences between the current and past research in error rates and the proportion of responses removed as RT outliers. For Study 1 reported here, the error rate was 4.3% and 4.4% of responses were removed in accordance with the outlier criterion. For past research (Bijlstra et al. 2010), there were 10.50% errors and <1% were removed as outliers. These effects are predictable - the 200 ms exposure duration likely encourages participants to respond quickly at the expense of accuracy leading to an overall reduction in RT latencies, (reducing the proportion of responses categorized as outliers) and increasing the number of errors.

Although direct comparison between the current and previous research is not possible because data for males were not reported separately in past research (and sample sizes were low), it is possible that reducing the exposure duration might affect performance in male participants specifically. As shown here, male participants were much slower to respond (See results for mu in Figure 5 and Figure 7), and ex-Gaussian analysis showed that the critical face sex X expression interaction effect for males was found across the RT distribution — the interaction effect was not restricted to the faster RTs as found for female participants. A practical implication of this observation is that encouraging participants to respond quickly by reducing exposure duration to 200 ms might make the pattern reported for male participants harder to detect because the full RT distribution is needed for the effects to emerge in males. In further support of this idea, the multiverse analysis showed that the 3-way face sex X expression X participant sex interaction was weaker (Bayes Factors were smaller) for outlier methods that removed large portions of the RT distribution. Considering these observations, I recommended that researchers use the exposure duration used both here and in separate research (e.g., Craig et al., 2018) namely, display faces until either a response has been made or 3-3.5 seconds has elapsed. Also, future studies should apply very lenient outlier removal (e.g., >150 ms and < 3500 ms) otherwise, it is possible that the outlier procedure will remove the effect of interest.

Strengths, Limitations, and Implications (Constraints on Generality)

A limitation of this study is that only sad and happy expressions were compared. Although past research has established that participants sex moderates the effect of face sex on angry vs happy expressions, it is yet to be established whether participant sex differences are also found for single valence conditions. For example, it has yet to be established whether participant sex effects are found for the comparison between fearful and sad expressions. If the current effects do indeed reflect evaluative associations triggered only in dual valance conditions, then the prediction will not be found for the comparison between fearful and sad expressions but rather there will be stereotype-based facilitation effects for female-sad (vs male-sad) and female-fearful (vs male-fearful) expressions irrespective of participant sex. Considering the differences in the distribution of the effects reported here, I recommend that researchers test such a hypothesis using either ex-Gaussian, ex-Wald or shifted-Wald analyses.

One strength of the current research is that it builds on previous research using the same large stimulus set used in past research and only a minor variation in method (participants pressed “S” for sad expressions rather than “A” for angry). The ex-Gaussian analysis is a strength because it includes an analysis of the time course of the effects of face sex and moreover, as a GLMM, the model includes a term to model by-item effects (stimulus identities). As outlined in several articles, by-item effects are important if researchers wish to generalize beyond the sample of stimuli used in the experiment. Finally, I have attempted to increase understanding by comparing effect sizes rather than the more usual emphasis on p-values.

The implications of this study are that participant sex is a critical variable that needs to be included in analyses of results of speeded expression recognition. Future studies are encouraged to look beyond non-significant p-values and report effect sizes for male and female participants separately using a model (e.g., the ex-Gaussian) that is well-suited to analyses of RT data. The groundwork for such an approach has already been provided in the current and recent work.

References

Becker, D. V., Kenrick, D. T., Neuberg, S. L., Blackwell, K. C., & Smith, D. M. (2007). The confounded nature of angry men and happy women. *Journal of Personality and Social Psychology*, *92*(2), 179–190. https://doi.org/10.1037/0022-3514.92.2.179

Bijlstra, G., Holland, R. W., & Wigboldus, D. H. J. (2010). The social face of emotion recognition: Evaluations versus stereotypes. *Journal of Experimental Social Psychology*, *46*(4), 657–663. https://doi.org/10.1016/j.jesp.2010.03.006

Brooks, J. A., Stolier, R. M., & Freeman, J. B. (2018). Stereotypes bias visual prototypes for sex and emotion categories. *Social Cognition*, *36*, 481–496. https://doi.org/10.1521/soco.2018.36.5.481

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, *77*(3), 305–327. https://doi.org/10.1111/j.2044-8295.1986.tb02199.x

Cousineau, D., & Chartier, S. (2010). Outliers detection and treatment: A review. *International Journal of Psychological Research*, *3*(1), Article 1. https://doi.org/10.21500/20112084.844

Craig, B. M., & Lipp, O. V. (2018). The influence of multiple social categories on emotion perception. *Journal of Experimental Social Psychology*, *75*, 27–35. https://doi.org/10.1016/j.jesp.2017.11.002

DeBruine, L. M., & Barr, D. J. (2021). Understanding Mixed-Effects Models Through Data Simulation. *Advances in Methods and Practices in Psychological Science*, *4*(1), 2515245920965119. https://doi.org/10.1177/2515245920965119

Eagly, A. H., & Mladinic, A. (1989). Gender Stereotypes and Attitudes Toward Women and Men. *Personality and Social Psychology Bulletin*, *15*(4), 543–558. https://doi.org/10.1177/0146167289154008

Ekman, P., & Friesen, W. V. (1976). *Pictures of facial affect*. Consulting Psychologists Press.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/bf03193146

Hugenberg, K., & Sczesny, S. (2006). On Wonderful Women and Seeing Smiles: Social Categorization Moderates the Happy Face Response Latency Advantage. *Social Cognition*, *24*(5), 516–539. https://doi.org/10.1521/soco.2006.24.5.516

Jeffreys, H. (1961). *Theory of probability.* (3d ed.). Clarendon Press.

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54–69. https://doi.org/10.1037/a0028347

Leppänen, J. M., & Hietanen, J. K. (2003). Affect and Face Perception: Odors Modulate the Recognition Advantage of Happy Faces. *Emotion*, *3*(4), 315–326.

Leppänen, J. M., Tenhunen, M., & Hietanen, J. K. (2003). Faster choice-reaction times to positive than to negative facial expressions: The role of cognitive and motor processes. *Journal of Psychophysiology*, *17*(3), 113–123. https://doi.org/10.1027/0269-8803.17.3.113

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

Lundqvist, D., Flykt, A., & Öhman, A. (1998). *The Karolinska Directed Emotional Faces*.

Mathot, S. (2017). *Bayes like a Baws: Interpreting Bayesian Repeated Measures in JASP*.

Nosek, B. A., & Banaji, M. R. (2001). The Go/No-go Association Task. *Social Cognition*, *19*(6), 625–666. https://doi.org/10.1521/soco.19.6.625.20886

Plant, E. A., Hyde, J. S., Keltner, D., & Devine, P. G. (2000). The Gender Stereotyping of Emotions. *Psychology of Women Quarterly*, *24*(1), 81–92. https://doi.org/10.1111/j.1471-6402.2000.tb01024.x

R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108. https://doi.org/10.1037/0033-295X.85.2.59

Richeson, J. A., & Ambady, N. (2001). Who’s in charge? Effects of situational roles on automatic gender bias. *Sex Roles: A Journal of Research*, *44*(9–10), 493–512. https://doi.org/10.1023/A:1012242123824

Rigby, R. A., & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape,(with discussion). *Applied Statistics*, *54*, 507–554.

Rudman, L. A., & Goodwin, S. A. (2004). Gender Differences in Automatic In-Group Bias: Why Do Women Like Women More Than Men Like Men? *Journal of Personality and Social Psychology*, *87*(4), 494–509. https://doi.org/10.1037/0022-3514.87.4.494

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, *110*, 104038. https://doi.org/10.1016/j.jml.2019.104038

Shackelford, T. K., Buss, D. M., & Weekes-Shackelford, V. A. (2003). Wife Killings Committed in the Context of a Lovers Triangle. *Basic and Applied Social Psychology*, *25*(2), 137–143. https://doi.org/10.1207/S15324834BASP2502_4

Smith, J. S., LaFrance, M., & Dovidio, J. F. (2017). Categorising intersectional targets: An "either/and" approach to race- and gender-emotion congruity. *Cognition & Emotion*, *31*(1), 83–97. https://doi.org/10.1080/02699931.2015.1081875

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing Transparency Through a Multiverse Analysis. *Perspectives on Psychological Science*, *11*(5), 702–712. https://doi.org/10.1177/1745691616658637

Taylor, S. E., Klein, L. C., Lewis, B. P., Gruenewald, T. L., Gurung, R. A. R., & Updegraff, J. A. (2000). Biobehavioral responses to stress in females: Tend-and-befriend, not fight-or-flight. *Psychological Review*, *107*, 411–429. https://doi.org/10.1037/0033-295X.107.3.411

Tipples, J. (2019). Recognising and reacting to angry and happy facial expressions: A diffusion model analysis. *Psychological Research*, *83*(1), 37–47. https://doi.org/10.1007/s00426-018-1092-6

Tipples, J. (2022a). Analyzing facial expression decision times: Reaction time distribution matters. *Emotion*, No Pagination Specified-No Pagination Specified. https://doi.org/10.1037/emo0001098

Tipples, J. (2022b). No need to collect more data: Ex-Gaussian modelling of existing data (Craig & Lipp, 2018) reveals an interactive effect of face race and face sex on speeded expression recognition. *Cognition and Emotion*, *0*(0), 1–8. https://doi.org/10.1080/02699931.2022.2120850

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

Voss, A., Voss, J., & Lerche, V. (2015). Assessing cognitive processes with diffusion model analyses: A tutorial based on fast-dm-30. *Frontiers in Psychology*, *6*. https://www.frontiersin.org/articles/10.3389/fpsyg.2015.00336