Supplementary Materials

Measuring the Impact of Multiple Social Cues to Advance Theory in Person Perception Research

Samuel A. W. Klein & Jeffrey W. Sherman

University of California, Davis

Public repository of data and analysis code: https://osf.io/gxbc5/

Table of Contents

Empirical Demonstration of the MCI Model 3
Method 3
Participants 3
Stimuli 3
Procedure 4
Analytic Approach 5
Results 11

Empirical Demonstration of the MCI Model

Method

Participants

In total, 619 undergraduates consented to participate for course credit. We decided a priori to exclude data from participants who performed at or below chance (errors on ≥50% of trials; n = 22) or who responded with the same key for more than 90% of trials (n = 2). The final sample comprised 593 participants (70.5% women; 45.5% Asian, 2.7% Black, 19.3% Latino, 18.2% White; M_{age} = 19.90, *SD* = 2.45).

Stimuli

This preregistered experiment relied on a stimulus set of faces selected from the Chicago Face Database (CFD; Ma et al., 2015). First, we sampled 10 female actors and 10 male actors from the CFD. When displaying neutral expressions, norming data form the CFD indicates that these actors' faces are correctly classified by gender >95% of the time. We then selected two face images per actor: one image in which they display a smile and another in which they display a scow. In total, we were left with 40 face images.

We then applied facial morphing techniques, using the online morphing tool *WebMorph* (Debruine, 2018), to manipulate ambiguity within the original stimulus set of 40 faces. Smiling and scowling face images, per actor, were blended together such that each actor was associated with two additional face images, besides their original smiling and scowling images: one image in which 55% of their smiling image and 45% of their scowling image remained (i.e., an ambiguous smile) and another in which 45% of their smiling image and 55% of their scowling image remained (i.e., an ambiguous scowl). At that point, we were left with double the original set of images – that is, 80 images in total. We then randomly paired male and female actors together, and morphed together their associated faces images. This secondary morphing procedure matched faces by their expression. If Female Actor A was randomly paired with Male Actor B, then their two unambiguous smiling images were morphed together, as were each of their other associated images (i.e., unambiguous scowling images, ambiguous smiling images, ambiguous scowling images). Morphs along sex generated ambiguous male (i.e., 45% male, 55% male) and ambiguous female (55% female, 45% male) face images. Because all image stimuli retained at least 55% properties of one level of sex and expression, each stimulus was associated with a correct response.

Procedure

In each experiment, participants were instructed to “quickly and accurately” classify a series of target faces. As previously mentioned, those targets varied in gender and emotion cues. Therefore, these two dimensions varied in relevance, depending on the task to which participants were assigned. Gender cues were considered task-relevant for those assigned to classify the faces by gender and task-irrelevant for those assigned to classify the faces by emotion. In contrast, emotion cues were considered task-relevant for those assigned to classify the faces by emotion and task-irrelevant for those assigned to classify the faces by gender.

On each trial, stimuli were presented in the following sequence: a fixation cross for 500 ms, a target face for 200 ms, and, finally, a pattern mask until a response was recorded. Following a block of 16 practice trials, participants completed three critical test blocks of 60, 60, and 40 trials, with self-paced breaks provided between blocks. Each face was presented just once during the critical test phase. Participants then provided demographics information before being debriefed on the purpose of the experiment. This experiment was approved by the University of California Institutional Review Board (protocol number: 223029; experiment name: *SHER806*).[1]

Analytic Approach

Heterogeneity of Variance. We first tested the heterogeneity of participant variance with an asymptotic chi-squared test which holds a null-hypothesis that response distributions are identical between participants. Indeed, there was a significant level of heterogeneity between participants' responses in both the gender classification condition, χ^2 (3564) = 6471.32, p < .001, and emotion classification condition, χ^2 (3528) = 6664.29, p < .001. Therefore, a hierarchical Bayesian approach (Klauer, 2010) was implemented using the *TreeBUGS* package (Heck et al., 2018) to estimate the MCI model parameters. This approach estimates individual-level parameter values under the assumption that each individual set of parameters follows the same hierarchical distribution, essentially treating participants as random factors.

Assessing Model Fit. Model fit, the ability for the model to predict observed response data, was mainly assessed via posterior predictive checks (PPCs). That is, we assessed the divergence between the observed response data and those predicted by the model's posterior distribution. Analytically, we conducted PPCs via two fit statistics. First, T_1 indicates the model's ability to account for discrepancies between observed and expected responses across all participants. That is, T_1 indicates mean-level fit. Second, T_2 indicates the model's ability to account for discrepancies between observed and expected covariances, providing an assessment of individual-level fit. We collected one thousand samples from the model's posterior distribution, and the means and covariances of the posterior were then compared to those in the observed data

[1] The preregistration is available at https://osf.io/m4eny.

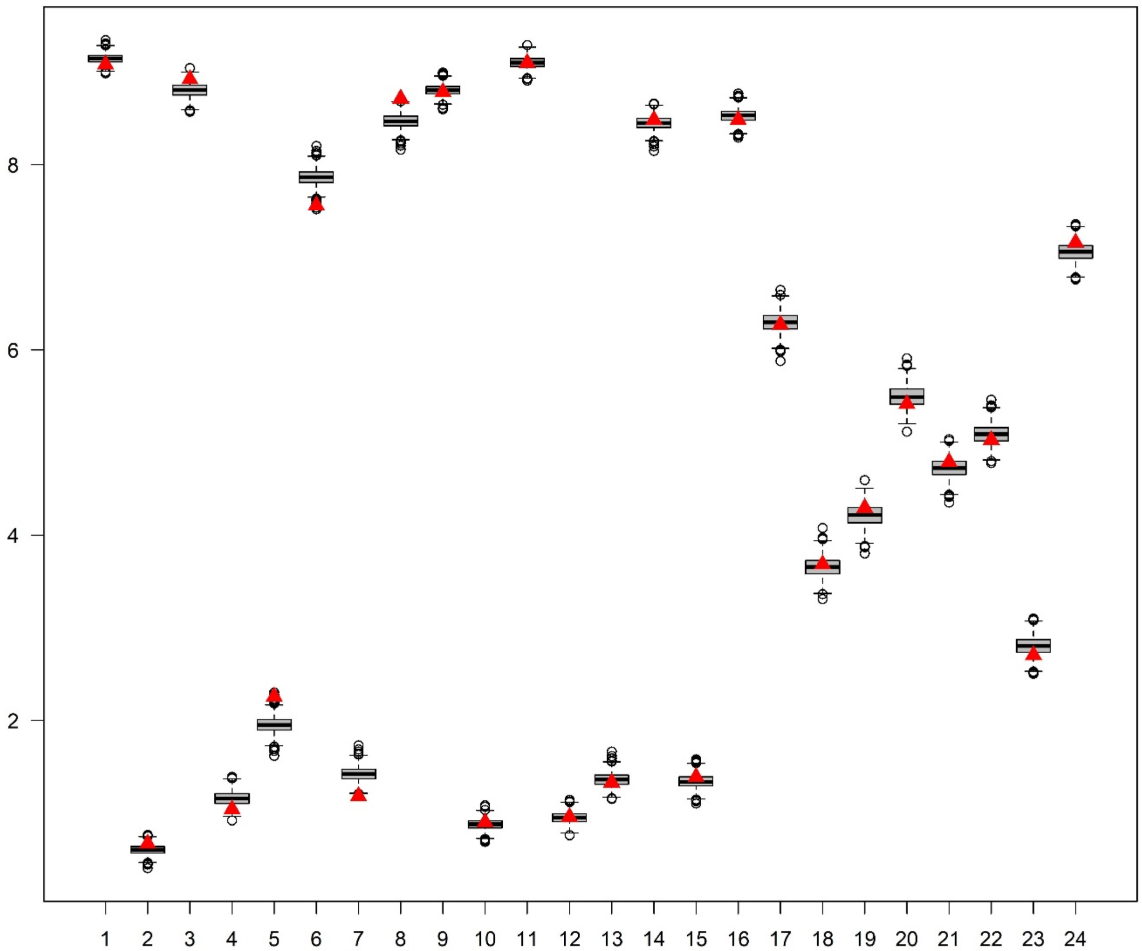
to produce T_1 (mean) and T_2 (covariance) statistics. Both T_1 and T_2 p-values reflect the probability that discrepancies between the expected and posterior-predicted data are greater than those between the expected and observed data (Heck et al., 2018, p. 273). The T_1 statistic is defined by a chi-square distribution and, therefore, must be considered in regards to the size of our dataset. The larger the dataset, the more sensitive the statistic will be at identifying even minor deviations of the model from the data. With N_{obs} ~70,000 observations, we expect a high level of sensitivity. Therefore, using the *w* statistic, we describe the level of misfit between the model's expectations and the observed aggregate responses. We rely on $w < .10$ as a descriptive benchmark for small levels of misfit.[2]

As stated in the main text, the MCI model fits well to both the gender classification judgments, Median Individual $T_{1\ p\text{-}value} = .558$, Aggregate $T_{1\ p\text{-}value} < .001$, Aggregate $T_{2\ p\text{-}value} = .002$, $w = .02$, and emotion classification judgments, Median Individual $T_{1\ p\text{-}value} = .538$, Aggregate $T_{1\ p\text{-}value} = .094$, Aggregate $T_{2\ p\text{-}value} = .192$, $w < .01$, albeit far better fitting for emotion classification judgments. Visually, we also plotted expected versus observed mean frequencies (T_1) and covariances (T_2). The observed means and covariances generally fall within the range of box-plotted model expectations, indicating good fit. Visual inspection is both inherently Bayesian and a common tool for MPT modeling using the current estimation strategy (e.g., Calanchini et al., 2021; Smith et al., 2019).

[2] Due to poor model fit when both sources of information were ambiguous, we removed the eight response categories from this condition before modeling the data. Although the preregistration states that we would model 32 response categories, and not 24, this decision avoids interpretations of a model that more poorly characterize the data.

Figure S1

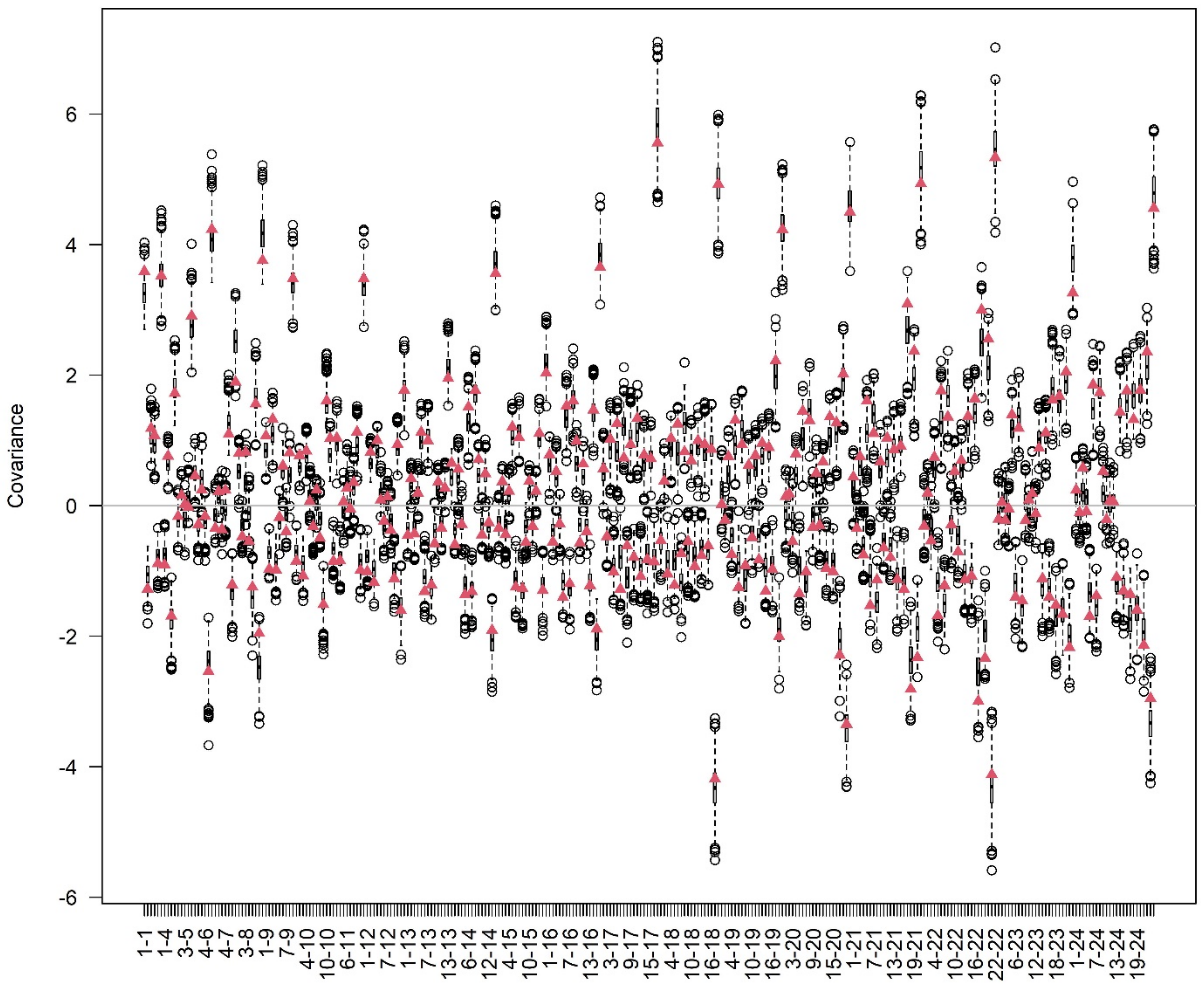
Plot of Posterior Predictive and Observed Mean Frequencies (Gender Classification)



Notes. Red triangle markers reflect observed mean frequencies. Boxplots reflect summarized frequencies from the fitted model's posterior distribution. Even-numbered response categories ("Woman" judgments) are the compliment of the odd-numbered response categories ("Man" judgments) and are, therefore, excluded from the plot.

Figure S2

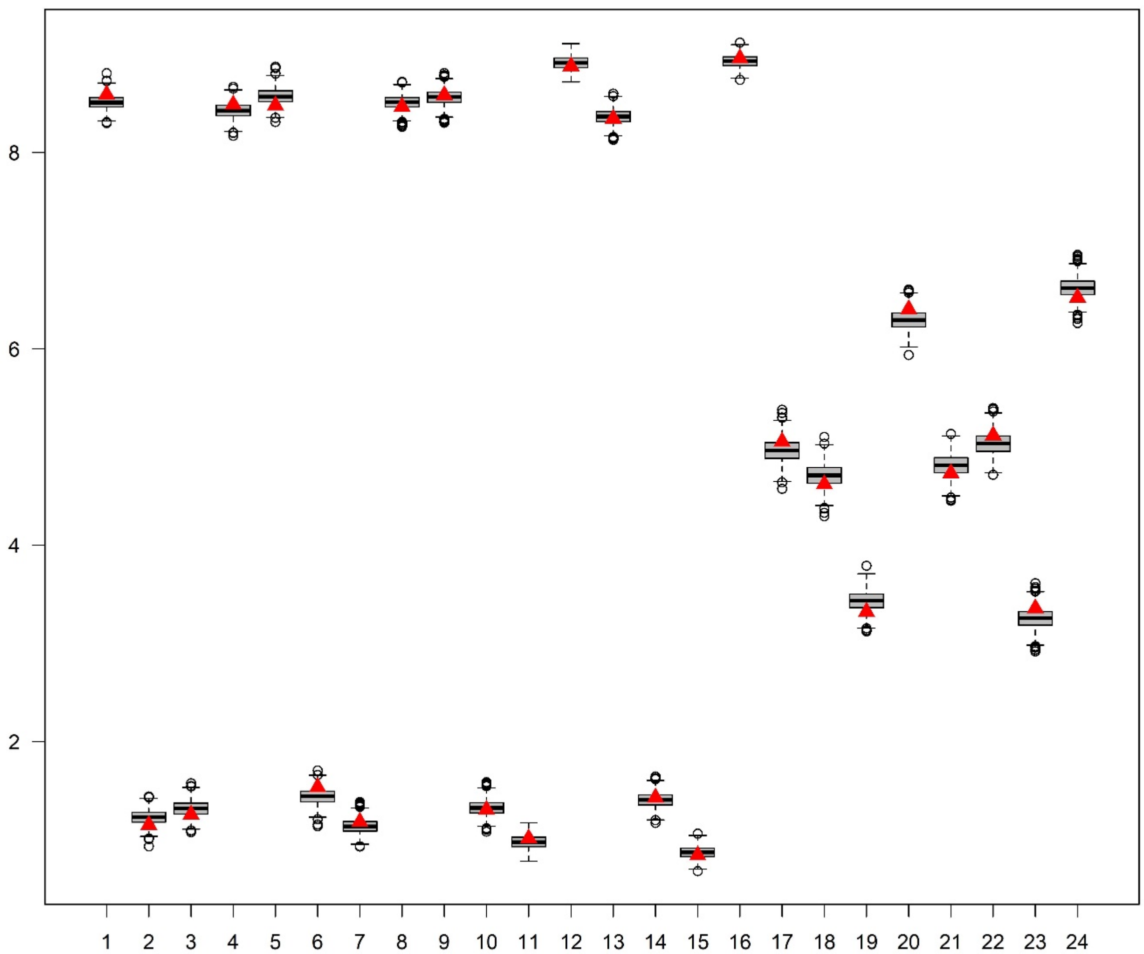
Plot of Posterior Predictive and Observed Covariances (Gender Classification)



Notes. Red triangle markers reflect observed covariances. Boxplots reflect summarized covariances sampled from the fitted model's posterior distribution. Even-numbered response categories ("Woman" judgments) are the compliment of the odd-numbered response categories ("Man" judgments) and are, therefore, excluded from the plot.

Figure S3

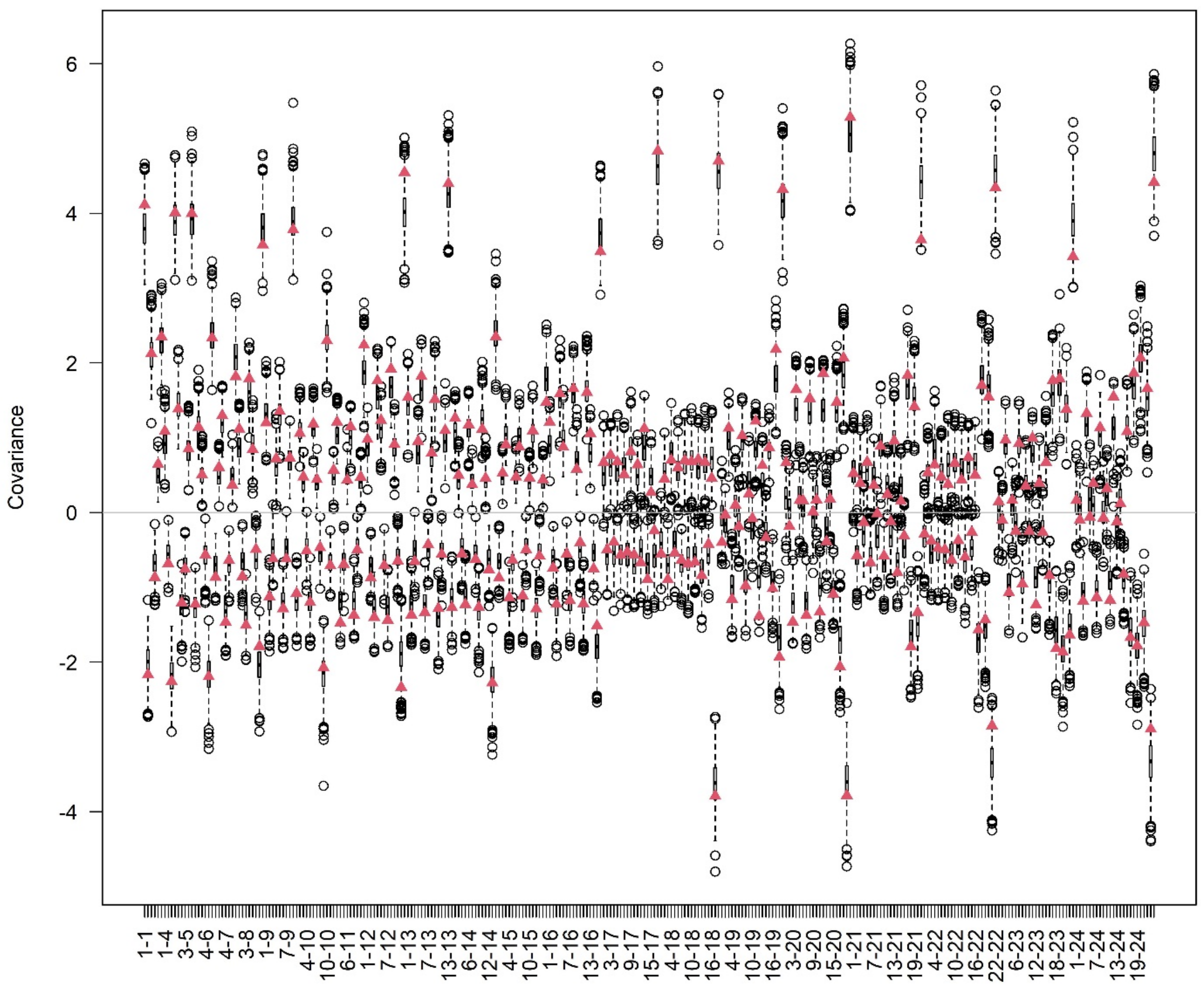
Plot of Posterior Predictive and Observed Mean Frequencies (Emotion Classification)



Notes. Red triangle markers reflect observed mean frequencies. Boxplots reflect summarized frequencies from the fitted model's posterior distribution. Even-numbered response categories ("Woman" judgments) are the compliment of the odd-numbered response categories ("Man" judgments) and are, therefore, excluded from the plot.

Figure S4

Plot of Posterior Predictive and Observed Covariances (Emotion Classification)



Notes. Red triangle markers reflect observed covariances. Boxplots reflect summarized covariances sampled from the fitted model's posterior distribution. Even-numbered response categories ("Woman" judgments) are the compliment of the odd-numbered response categories ("Man" judgments) and are, therefore, excluded from the plot.

Results

Gender Classification. Parameter comparisons revealed a credible and large effect of ambiguity in gender cues on gender processing, $\Delta C_1 = .62$, 95% BCI [.59, .64], indicating that the use of gender cues to classify faces by gender decreased with ambiguity. Ambiguity in emotion cues also decreased the use of gender cues to classify faces by gender, $\Delta C_1 = -.04$, 95% BCI [-.06, -.02], although this effect was >30 times smaller than the effect of ambiguity in gender cues on gender processing, $\Delta C_1 / \Delta C_1 = 32$.

Ambiguity in emotion cues also had a credible effect on emotion processing, $\Delta C_2 = .23$, 95% BCI [.16, .29], indicating that the use of emotion cues to classify faces by gender decreased when they were ambiguous. The effect of emotion cue ambiguity on emotion processing was >5 times larger than its effect on gender processing, $\Delta C_2/\Delta C_1 = 5.75$.

Emotion Classification. Parameter comparisons revealed a credible and large effect of Ambiguity on emotion processing, $\Delta C_2 = .64$, 95% BCI [.60, .67], indicating that the use of emotion cues to classify faces by emotion decreased with ambiguity. Ambiguity in gender cues also decreased the use of emotion cues to classify faces by emotion, $\Delta C_2 = -.03$, 95% BCI [-.05, >-.01], although this effect was >25 times smaller than the effect of emotion cue ambiguity on emotion processing, $\Delta C_2 / \Delta C_2 = 25.22$.

Task Relevance. To explore whether each cue is used more if it is relevant versus irrelevant to the intended judgment, we compared parameter estimates between gender and emotion classification tasks. Gender cues were used more when they were task-relevant versus task-irrelevant, $\Delta C_1 \geq .13$, 95% BCI [≥.11, ≥.15]. Emotion cues were used more often when they were task-relevant versus task-irrelevant, $\Delta C_2 \geq .14$, 95% BCI [≥.11, ≥.17] (see Table S1).

Table S1

Analyses of C1 and C2 Parameters Across Conditions of Cue Ambiguity

Effect	*Mean Difference*	*SD*	95% *BCI*
Use of Gender Cues (C_1)			
Unambiguous Gender and Emotion Cues	.74	.01	[.72, .77]
Ambiguous Gender Cues	.13	.01	[.11, .15]
Ambiguous Emotion Cues	.78	.01	[.75, .81]
Use of Emotion Cues (C_2)			
Unambiguous Gender and Emotion Cues	.55	.04	[.48, .62]
Ambiguous Gender Cues	.59	.02	[.56, .63]
Ambiguous Emotion Cues	.14	.02	[.11, .17]