Title
Modeling Cue-integration in Emotion Inferences

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 44(44)

Authors
Goel, Srishti
Jara-Ettinger, Julian
Gendron, Maria T

Publication Date
2022

Peer reviewed

Modeling Cue-integration in Emotion Inferences

Srishti Goel, Julian Jara-Ettinger, Maria Gendron
(srishti.goel@yale.edu, julian.jara-ettinger@yale.edu, maria.gendron@yale.edu)
Department of Psychology, 2 Hillhouse Ave.
New Haven, CT 06511, USA

Abstract

Inferences of other's emotion states are influenced by multiple sources including target cues such as facial movements and the situational context. Our understanding of how information from these cues is integrated is limited, however. We examined whether people integrate information from faces and situations to infer emotions as predicted by an existing model of affective cognition. We applied a Bayesian cue-integration model to a dataset that includes a variety of complex social situations that reflect the heterogeneity of emotion contexts in social lives. Results indicate that when viewing both faces and situations, situation information alone predicted people's inference about emotions better than Bayesian cue-integration model. However, there was some variability in this pattern across emotion categories as the Bayesian cue-integration model best predicted inferences for emotion categories of *amusement* and *happiness*. These findings better our understanding of the interplay between facial and situational cues in informing emotion inferences.

Keywords: Emotion inference, Bayesian modeling, Affective cognition

Introduction

People routinely make inferences about other's emotional states, with profound consequences for our interpersonal functioning, including how well we navigate social relationships (for meta-analytic review, see Hall, Andrzejewski, & Yopchick, 2009). Developing a robust scientific account of how we make these inferences is important. Commonsense experiences suggest that we make emotion inferences with relative ease. Some scientific evidence aligns with this alacrity, including the efficiency with which we process facial cues of emotion (for review, see Spunt & Adolphs, 2019). Despite the disproportionate focus of emotion research on how people process isolated canonical facial portrayals (or, facial expressions) (Gendron & Barrett, 2017), accumulating evidence suggests emotion perception is more complex (Barrett, Mesquita, & Gendron, 2011; Gendron, Mesquita, & Barrett, 2013; Hareli, Elkabetz, & Hess, 2019). For example, target cues like facial movements rarely conform to the canonical expressions used in scientific studies during instances of emotion (correlation ranges between 0.13 and 0.30; see, Durán & Fernández-Dols, 2021; Barrett, et al., 2019). Further, evidence suggests that perceivers heavily rely on context to make inferences of emotion (e.g., Le Mau, et al., 2021; for review see, Barrett, et al., 2011; Aviezer, Ensenberg, & Hassin, 2017).

One contextual source widely examined in emotion perception is knowledge of the situational context, which is often social in nature (for e.g., Carroll & Russell, 1996). Situational context is often operationalized as description of emotionally laden events that elicit an emotional response, as it is in this report. Early research examining the role of situational context and facial portrayals was often aimed at establishing the dominance of one cue over the other in inferring emotions (for e.g., Wallbott, 1998; Carroll & Russell, 1996). This often involved using canonical facial cues and situational cues that were "clear" and evocative (see, Ekman, Freisen, & Ellsworth, 1972, for critiques on clarity of stimuli). Meaning, that each cue independently elicits a high consensus agreement on a particular emotion. The use of such stimuli can compromise ecological validity as canonical expressions are rare and often not correlated highly with emotional experiences as predicted (Durán & Fernández-Dols, 2021). Instead, people often make inferences based on ambiguous, complex, and subtly expressive facial and contextual information (for review, Barrett et al., 2019). An open question is how large of a role the face plays in inferences under these circumstances. Additionally, situations can carry added complexity because people can experience a range of emotions in a particular situation (for discussion see, Hoemann, Gendron, & Barrett, 2017). This leaves open the question of how individuals integrate situational information into emotional inferences. In the present research, we address these issues by extending a previously proposed model of affective cognition to a data set that includes greater complexity, and diversity of emotional experiences.

Bayesian Model for Affective Cognition

Computational models of cognition have been increasingly used to formalize and test human lay theories. One particularly fruitful approach, known as *rational analysis* (Anderson, 1996), aims to understand human reasoning by comparing human judgments against the predictions of normative models that capture different hypotheses about the computations in people's minds (for affective cognition, see Ong, Zaki, & Goodman, 2019; for social cognition see,

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society*.

Baker, Jara-Ettinger, Saxe, & Tenenbaum, 2017). Despite their widespread application in cognition for over a decade, Bayesian computational models have only recently been applied to studying the cognitive processes involved in reasoning about emotions, also referred to as affective cognition.

Recently, researchers have proposed that humans understand each other's emotions though a *lay theory of emotion* (Ong, Zaki, & Goodman, 2015). This theory captures our expectations of how situational outcomes cause emotions, and how emotions then cause facial expressions. Equipped with this causal model, people can then use Bayesian inference to determine what emotions people are likely experiencing, given information about their facial expressions and the situation that they're in. Ong and colleagues (2015) presented a formal computational model of this idea, which posits that inferences about emotions follow a form of cue integration given by:

$$\boldsymbol{P(e|o,f) \propto \frac{P(e|o)P(e|f)}{P(e)}} \quad \textbf{(1)}$$

Here, $P(e|o,f)$, represents observers' beliefs that an agent is experiencing emotion *e*, given their facial expression *f* and an outcome *o*. If people infer each other's emotions by applying Bayesian inference to a lay theory of emotion, then people's inferences should be proportional to the product of likelihood of each individual cue $P(e|o)$ (the probability of emotion *e* given outcome *o*) and $P(e|f)$ (the probability of emotion *e* given facial expression *f*), divided by the prior probability of an emotion occurring $P(e)$.

Ong and colleagues (2015) demonstrated that their cue-integration model closely tracked the empirical data – i.e., people's judgments of emotions when they were asked to make inferences given both the facial expressions and situational outcomes. However, their experiment tested the model on a dataset consisting of a narrow set of situations: different outcomes within a gambling game; their design restricted situational outcomes to the amount of money won from spinning a wheel with different winning probabilities. Additionally, they computed an overall metric of model fit across different emotion categories but did not examine variability in model performance for different emotion categories (e.g., *anger*, *fear*, *happiness*, etc.) Although the structure of this experiment was useful as a first test of their account, this work leaves open the questions of whether the same computations can capture emotion inferences in a broader range of situations that include a diverse set of realistic emotional experiences, and whether these computations similarly capture inferences for different emotion categories.

In this project, we examined the robustness of this modeling approach by applying it to this relatively more ecologically valid dataset (described in more detail in the next section) and examining it for each emotion category separately. We also propose that by utilizing naturalistic and realistic portrayals of facial information, we can better model how emotion inference unfolds in the real world.

The Present Study

In the current research, we had two primary aims. First, we aimed to examine the robustness of the Bayesian model for cue-integration $P(e|o,f)$ (Ong, et al., 2015) using an archival data set (Le Mau et al., 2021) and an empirically collected data set on emotion priors $P(e)$. Specifically, this involved examining the cue-integration model in comparison with two alternative models each of which rely on a single cue – Face-only $P(e|f)$ and Situation-only $P(e|o)$ respectively. We hypothesized that the cue-integration model might not fully capture emotion inference in light of a more diverse set of complex situations than those explored in prior work (e.g., winning or losing in a game). Specifically, our research question was whether the Bayesian cue-integration model would track with judgements of emotion when the cues were more naturalistic and, in many cases, high in ambiguity.

Second, we aimed to examine the applicability of the cue-integration model to specific emotion categories (e.g., anger, fear, etc.). Ong and colleagues (2015) computed model fit by correlating the model-based results with empirical ratings across all emotions categories. Here, we aimed to examine the robustness of cue-integration model for each emotion category separately. We hypothesized that there would be variability across emotion categories in how closely the cue-integration model versus single cue models track people's empirical judgements. Not only do people report seeing expressions of emotions in varying amounts in everyday life (Somerville & Whalen, 2006), but predicted expressions (e.g., widened eyes, gasping in fear) also vary in reliability with actual experience (Duran & Fernandez-Dols, 2021). These findings suggest that cue-reliance on the face may be lower for emotional experiences with less reliable facial signals.

To generate the cue-integration model predictions, it is necessary to obtain estimates of $P(e|o)$, $P(e|f)$, and $P(e)$, for all emotions, facial expressions, and event outcomes that we consider (see Equation 1). In our task, the first two terms were obtained from archival data from Le Mau et al., (2021), and the final term was obtained via an Emotion prior estimation task.

In addition to the cue-integration model based on Ong et al., (2015), which we call the Bayes cue-integration model, we also considered two simpler alternative models. The Situation-only model $P(e|o)$ which represents beliefs that an agent experiences emotions due to situation outcomes, or as we call situations, and the Face-only model $P(e|f)$ which represents beliefs that an agent's experience of emotions leads them to express it in terms of external cues like facial movements. Predictions of these single cue models were obtained from the archival data from Le Mau et al., (2021) by computing probability estimates for inferred emotions when participants viewed situation descriptions $P(e|o)$ alone and when they viewed facial portrayals $P(e|f)$ alone. These simpler models suggest that an observer would rely on a single cue (face or situation) to approximate their inference of emotions when presented with both faces and situations. Finally, predictions from each of the three models were compared to participants' judgements obtained from the archival dataset from Le Mau et al., (2021) i.e., participants' inferences of emotions when they saw both the situational description and facial portrayals, which we call empirical cue-integration estimates.

Archival Data

The archival data (Le Mau., et al., 2021) contains ratings of social scenarios, ratings of actors' portrayals of those scenarios, and ratings of the combination. These ratings were obtained from a total of 604 stimuli pairs (scenarios and poses) that were sourced from two books: *In Character: Actors Acting* (Schatz, 2006) and *Caught in the Act: Actors Acting* (Schatz, 2013). These volumes contain images of expressions posed by a pool of professional actors[1] after they were provided with emotionally evocative scenarios. A couple example of the scenarios are – 'She is confronting her lover, who has rejected her, and his wife as they come out of a restaurant', 'He is a motorcycle dude coming out of a biker bar just as a guy in a Porsche backs into his gleaming Harley'.

Actors' facial portrayals did not reliably align with proposed canonical facial configurations of emotion categories (Le Mau, et al., 2021). Instead, the facial poses conformed with the variability observed in spontaneous emotion expressions in everyday life. For example, similar to people in their daily lives, actors scowled about 30% of the times when portraying scenarios consistent with the emotion anger. Given the greater complexity of this stimulus set, including the scenarios and the portrayals of them, this stimulus set offers greater ecological validity and range.

The data included 75390 observations by participants who rated a random subset of approximately 30 stimuli (out of a total of 604 stimuli). Each observation included ratings on 13 different emotion categories. Participants provided these ratings for one in 3 different conditions - face-only (*N*=842), situation-only (*N*=839), face and situation combined (*N*=845). In the face-only condition participants viewed only the actor's portrayals of scenarios. In the situation-only condition, participants viewed only the description of those scenarios. In the combined condition, participants viewed both the description of scenarios along with the actor's portrayals of those scenarios. Participants were asked to provide ratings only in one of the three conditions. The 13 emotions they rated were - *amusement, anger, awe, contempt, disgust, embarrassment, fear, happiness, interest, pride, sadness, shame, surprise*. Each emotion category was first rated for presence, i.e., if the participant observed an emotion they respond saying 'yes' or 'no'. If the emotion was present, i.e., they responded yes, then the participant rated the intensity of that emotion on a 4-point-likert type scale (slightly, moderately, strongly, intensely) (Russell & Carroll, 1999).

Emotion Prior Estimation Task

To calculate model predictions, it is necessary to compute people's prior probabilities over emotions ($P(e)$). We therefore collected rating data for people's likelihood of perceiving each of the 13 emotion categories in their everyday lives (*prior task*). We collected empirical priors rather than estimating priors based on the archival data directly because we cannot assume the range of stimuli is representative of base rates of emotional instances in everyday life that likely inform priors. We also included a short *rating task* that was identical to the combined condition from Le Mau and colleagues' paper (2021) to ensure consistency between our collected data and the archival data i.e., to examine whether providing such ratings would change the nature of judgements.[2] We also compared these informative emotion priors to a flat prior to examine whether informative priors improve model fit over a flat prior.

Participants We recruited 45 native English-speaking participants from the US (20 male, 25 female, mean age = 38-year, age range = 18-60 years) using online data collection platform Prolific. The number of participants was determined based on the average number of participants responding to a stimulus in the archival data.

Stimuli The stimuli consisted of a question that asked people's likelihood of perceiving each emotion category in their daily lives on a 7-point-likert type scale (1: not at all likely, 4: moderately likely, 7: extremely likely). The question stated: 'When you see people experiencing emotions in your day-to-day life, how likely are you to perceive people experiencing {Emotion}'. Empirically collecting priors by asking participants to rate a single question is a standard procedure in the literature (for example, Wu, et al., 2018) and these priors were compared to similar ratings collected by Sommerville & Whalen (2005). Additionally, a random sample of 10 stimuli was drawn from the larger pool of 604 stimuli in the archival data for collection of joint-cue ratings.

Procedure All participants read an online consent form before agreeing to participate in the Study. Participants then read instructions for the task where they would provide ratings for emotions after viewing descriptions of scenarios and actor's portrayals of emotions in those scenarios. The instructions were identical to those provided by Le Mau and colleagues (2021). This was followed by a question aimed to validate that they read the instructions. After answering this question, participants were asked to give an estimate of how likely they think certain emotions occur, before proceeding

[1] We are unable to provide example images of the facial portrayals because the images are under copyright.

[2] We confirmed that the distribution of ratings for the subset of 10 stimuli were similar in the *rating task* and the archival data for all 13 emotion categories.

to the *rating task*. They read instructions for providing rating on the likelihood of perceiving each emotion category and responded to another question to validate that they read the instructions. Next, they provided ratings on the *prior task* followed by the brief 10-item *rating task*. Lastly, participants also filled out a brief demographic questionnaire at the end.

Results

Overall Model Analysis

We first computed the empirical cue-integration probabilities by averaging the ratings for each stimulus on the joint-cue trials i.e., trials where both face and situation combinations were present (using the archival data). We then computed the Face-only $P(e|f)$ and Situation-only model $P(e|s)$ probabilities by again averaging the ratings for each stimulus in the Face-only and Situation-only conditions respectively (again using the archival data) (see distribution of raw ratings for a particular stimulus, Fig. 1)[3]. Finally, the face-only and Situation-only model probabilities along with the emotion prior probabilities $P(e)$ were used in equation 1 to compute the Bayes cue-integration model probabilities $P(e|f,s)$. All model and empirical probabilities were normalized such that the sum of probabilities for rating the 13 emotions for a particular stimuli is equal to 1.

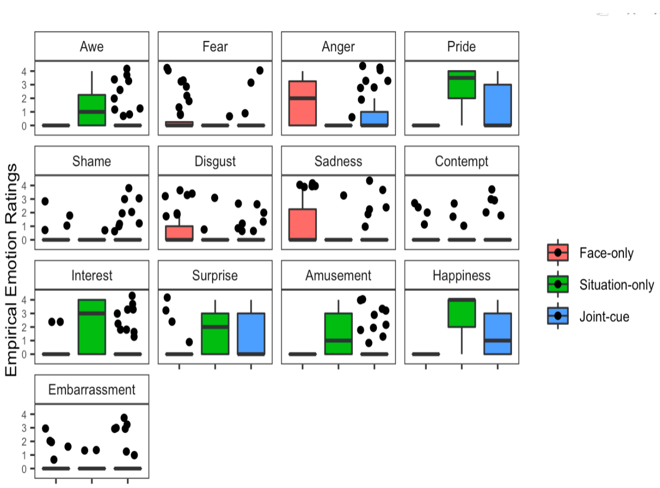

Figure 1: Distribution of participants' ratings for Stimulus #362 faceted by emotion. From left, boxplot in each graph represents ratings from Face-only condition (red), Situation-only condition (green), and Joint-cue condition (blue) respectively.

To compare model predictions to participant's judgements, we correlated the empirical cue-integration probabilities with each of the model probabilities (Fig. 2). Participant's empirical judgements had a significantly strong correlation with all models. The Situation-only model had the highest correlation with a value of $r = 0.86$ ($p < 0.0001$) and a 95% confidence interval (0.85, 0.87), followed by the Bayes cue-integration model with a value of $r = 0.84$ ($p < 0.0001$) and 95% confidence interval (0.83, 0.85), and finally the Face-only model with a value of $r = 0.66$ ($p < 0.0001$) and 95% confidence interval (0.64,0.68).

Our results suggest that the situation-only model best tracked people's inference of emotions in presence of both facial and situational cues. There was a statistically significant difference in the Situation-only and Bayes cue-integration model correlations ($t = -7.19$, $p < 0.0001$) as well as the Situation-only and Face-only model ($t = 36.31$, $p < 0.0001$) based on tests of significance for a difference between two dependent correlations.

We also computed the Bayes-cue integration model using a flat prior to examine whether empirically collected priors are informative and improve model fit over a flat prior. The Bayes cue-integration model with flat priors correlates with empirical data ($r = 0.83$, $p < 0.0001$) significantly less compared to the model with informative priors ($t = 9.13$, $p < 0.0001$). This suggests that global priors of emotions are informative and add value to the cue-integration model.

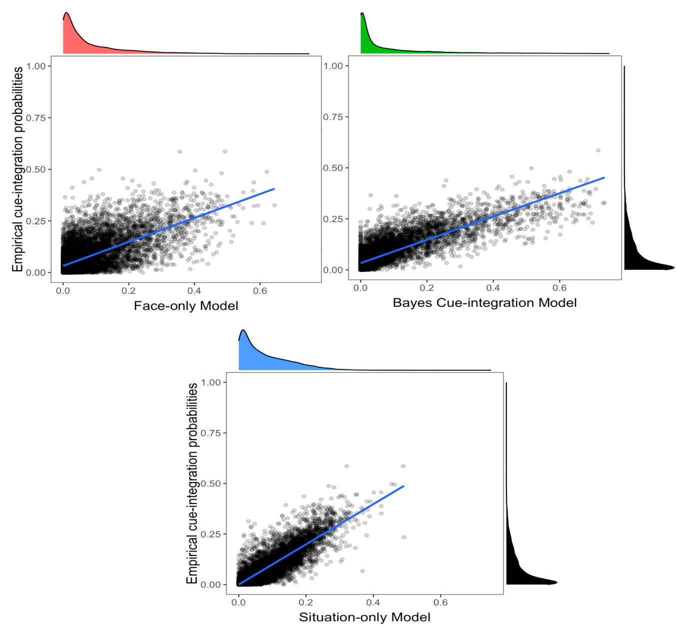

Figure 2: Overall model correlations for Face-only, Bayes Cue-integration, and Situation-only models respectively. Each point represents the probability of inferring a particular emotion for a stimulus. The blue line shows the best linear fit between the model and the data. Density distributions of model probabilities and empirical judgements are provided next to the respective axis on each sub-plot.

[3] We do not compute probabilities using kernel density estimates (KDE), as done in previous work, because we are estimating inference of a categorical variable (emotion), not a continuous variable. KDE can estimate probability distribution for rating values (e.g., 1,2) on an emotion and stimulus but the outcome of interest here is probability distribution for inferring emotions given a stimulus.

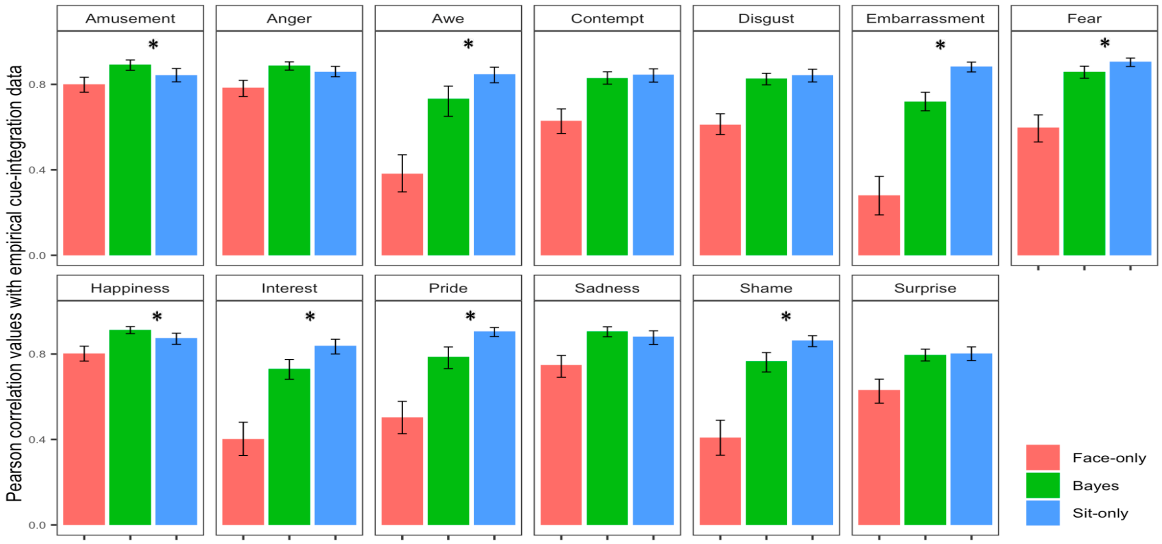

Figure 3: Pearson correlation of the models with the empirical judgements from joint-cue condition faceted by each emotion category. The error bars represent the bootstrapped 95% confidence intervals. The Face-only model correlation is significantly lower than Bayes cue-integration and Situation-only model (except *Amusement*) for all emotion categories. Asterisk (*) denotes significant difference between Bayes cue-integration and Situation-only model correlations.

We also computed root-mean-squared-error (RMSE) to assess model predictions. RMSE estimates corroborate with the correlations. Situation-only model had the lowest RMSE value (0.042, bootstrapped 95% confidence interval: 0.041, 0.044), followed by the Bayes cue-integration model (0.071, bootstrapped 95% confidence interval: 0.068, 0.073), and finally the Face-only model (0.075, bootstrapped 95% confidence interval: 0.073, 0.077). This suggests that people's inference of emotions from just the situational cues best predicted their inference of emotions when they had access to both facial and situational cues. All code and data are available at - https://osf.io/j79sy/.

Emotion specific Analysis

The overall pattern of results was largely consistent across emotion categories (Fig. 3). The Situation-only model and the Bayes cue-integration model systematically performed better than the Face-only model. However, as predicted, there was some variability across emotions. Four emotion categories differed from this overall pattern as the highest correlation value was observed for Bayes cue-integration model instead of the Situation-only model – *Amusement* (Bayes cue-integration $r =$ 0.89, Situation-only r = 0.84), *Anger* (Bayes cue-integration $r =$ 0.89, Situation-only r = 0.86), *Happiness* (Bayes cue-integration $r =$ 0.91, Situation-only r = 0.87), and *Sadness* (Bayes cue-integration $r =$ 0.90, Situation-only r = 0.88). Confidence intervals for bootstrapped samples of difference in correlation values between all three models were computed for all emotion categories. Compared to the Bayes-cue integration model, the Situation-only model correlation was significantly higher for *Awe, Embarrassment, Fear, Interest, Pride, Shame*; was significantly lower for *Amusement, Happiness*; and was not statistically different from the Bayes cue-integration model for *Anger, Contempt, Disgust, Sadness, Surprise*. This finding suggests that for emotion categories of *Amusement* and *Happiness*, inferences from the Bayes cue-integration model tracked people's empirical judgments more closely than the Situation-only model. For emotion categories of *Anger, Contempt, Disgust, Sadness* and *Surprise*, both the Bayes cue-integration and Situation-only tracked people's empirical judgments closely. All code and data are available at - https://osf.io/j79sy/.

To further explore this variability, we computed a diagnosticity metric for each emotion category by computing the percentage of stimuli where the emotion was perceived as strong to intense i.e., where the average participant rating was above 2.5 (on a 0-4 rating scale) (Table 1). This operationalization of diagnosticity is consistent with earlier work examining role of diagnosticity of information in trait-based inferences (for example, Cone & Ferguson, 2015). Based on this diagnosticity variable, situational information appears to be overall more diagnostic for all emotion categories. However, there is variability in the perceived diagnosticity of faces and situations across emotion categories. Emotion diagnosticity patterns largely align with which emotion categories evince greater integration between face and context, with the exception of surprise. Future work is necessary to manipulate diagnosticity and/or the degree to which facial expressions conform to canonical patterns to examine the impact on integration.

It is also important to take note of a caveat; this diagnosticity variable treats high-intensity expressions as most diagnostic. It is possible, however, that expressions can be diagnostic, but for low intensity expressions (this would be evidenced by a narrow rating distribution that is relatively low on the intensity scale). Future work should explore additional ways of computing diagnosticity such that participant's consensus on ratings and intensity of emotion perceived are not confounded. For example, computing a metric by weighing the intensity and variance of ratings.

Table 1: Percentage of facial and situational stimuli perceived diagnostic for each emotion category.

Emotion	% facial stimuli perceived diagnostic	% situational stimuli perceived diagnostic
Amusement	1.99	1.99
Awe	0	1.49
Anger	5.96	17.55
Contempt	0	1.16
Disgust	0.45	6.29
Embarrassment	0	6.95
Fear	0.83	13.41
Happiness	9.27	10.6
Interest	0	7.28
Pride	0	5.63
Sadness	1.99	12.75
Shame	0	5.3
Surprise	4.47	14.9

Discussion

In the present report, we extended the application of a Bayesian model for cue-integration in affective cognition (Ong, et. al., 2015) to a more complex, diverse, and arguably, ecologically valid archival dataset (Le Mau, et. al., 2021) to test the model's robustness in capturing integration of information from facial and situational cues in making inferences about other's emotional states. We also examined the robustness of this model separately for different emotion categories.

Our results suggested that overall people's inferences about other's emotions based on access to both facial and situational cues (joint-cue condition) were best predicted by the situational information alone (Situation-only model), when evaluating model fit. The Bayesian cue-integration model also closely tracked with people's inferences of emotions but under-performed in comparison to the Situation-only model. Lastly, people's inferences of emotions from both cues were least predicted by the facial information alone (Face-only model). These findings suggest that inferences of emotions appear to rely more heavily on situational information.

When we examined model fit by emotion, there were some emotions (happiness, amusement) where Bayesian cue-integration best accounted for participants inferences. The emotions that individuals indicate seeing most often in people's facial behaviors (Somerville & Whalen, 2006) were also the emotions where we observe more integration of the facial cues (rather than inferences driven by the situation alone). We speculate that variation in how often real-world social signals conform to similar expressive configurations impacts the mental representations of facial expressions for certain emotion categories that perceivers use to guide their inferences. Additional research is warranted to examine these emotion-specific patterns in integration in other stimuli and contexts.

A strong reliance on situational context is consistent with past literature examining inferences from spontaneous expressions derived from real-world contexts. For example, spontaneous real-world facial movements in high intensity contexts of sporting wins and losses have no utility in discriminating valence-based information beyond bodily cues (Aviezer, Trope, & Todorov, 2012). Our findings build on these to suggest that non-canonical facial portrayals capturing a broader range of intensities and emotions are similarly less informative for emotion inferences than the context (conceptually replicating Le Mau et al., 2021). This may be influenced by the low diagnosticity of the facial portrayals, which are consistent with recent evidence that people rarely generate canonical faces that are proposed in the literature (Durán & Fernández-Dols, 2021).

Our findings also parallel the literature examining the role of facial information in impression formation and updating. Research on trait inferences suggests that people infer traits from appearance-based cues from faces (e.g., Todorov & Ducahine, 2008) especially when asked to make speeded judgements (e.g., Willis & Todorov, 2006; Blair, Chapleau, & Judd, 2005). However, both explicit (Rule, Tskhay, Freeman, & Ambady, 2014) and implicit (Shen, Mann, & Ferguson, 2020) impressions of traits based on facial information are updated in presence of explicit descriptions of target behaviors/characteristics. Together with our results, findings such as these suggest that the assumption of facial dominance in social cognition may be overstated. People's inferences of others, be that trait or state inferences, can be formulated based on facial information as countess studies have shown, but appear to be heavily constrained by other sources of information such as knowledge of the context. Indeed, prior research suggests that perceivers overestimate the value of facial information over context in forming inferences about emotion (Zhou, Majka, & Epley, 2017).

Future research unpacking the situational and facial information will be crucial to understand the boundary conditions of these findings. Further, investigating how greater degrees of uncertainty in inferences formed based on facial and situational information affects integration is an important open question. For example, many scenarios take on a more omniscient perspective and provide privileged knowledge that may inflate certainty of inferences beyond what is typical in daily life. Investigating the effects of titrating information on perceiver's inferences would be an interesting future direction.

References

Anderson, J.R. (1996). ACT: A simple theory of complex cognition. American psychologist, 51(4), 355.

Aviezer, H., Ensenberg, N., & Hassin, R. R. (2017). The inherently contextualized nature of facial emotion perception. *Current opinion in psychology*, *17*, 47-54.

Aviezer, H., Trope, Y., & Todorov, A. (2012). Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science*, *338*(6111), 1225-1229.

Baker, C. L., Jara-Ettinger, J., Saxe, R., & Tenenbaum, J. B. (2017). Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behavior*, *1*(4), 1-10.

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological science in the public interest*, *20*(1), 1-68.

Barrett, L. F., Mesquita, B., & Gendron, M. (2011). Context in emotion perception. *Current Directions in Psychological Science*, *20*(5), 286-290.

Blair, I. V., Chapleau, K. M., & Judd, C. M. (2005). The use of Afrocentric features as cues for judgment in the presence of diagnostic information. *European Journal of Social Psychology*, *35*(1), 59-68.

Carroll J. M., & Russell, J. A. (1996). Do facial expression signal specific emotions? Judging emotion from the face in context. Journal of Personality and Social Psychology, 70, 205–218.

Cone, J., & Ferguson, M. J. (2015). He did what? The role of diagnosticity in revising implicit evaluations. *Journal of Personality and Social Psychology*, *108*(1), 37.

Durán, J. I., & Fernández-Dols, J. M. (2021). Do emotions result in their predicted facial expressions? A meta-analysis of studies on the co-occurrence of expression and emotion. *Emotion*.

Ekman P. (1972). Universal and cultural differences in facial expressions of emotions. In J. K. Cole (Ed.), *Nebraska symposium on motivation*, *1971* (pp. 207–283). Lincoln: University of Nebraska Press.

Ekman, P. (1982). What are the relative contributions of facial behavior and contextual information to the judgement of emotion?. *Emotion in the human face*, 111-127.

Ekman P., Freisen W. V., & Ellsworth P. (1972). Emotion in the human face. New York: Pergamon Press.

Gendron, M., & Barrett, L. F. (2017). Facing the past. *The science of facial expression*.

Gendron, M., Mesquita, B., & Barrett, L. F. (2013). Emotion perception: Putting the face in context.

Hall, J. A., Andrzejewski, S. A., & Yopchick, J. E. (2009). Psychosocial correlates of interpersonal sensitivity: A meta-analysis. *Journal of nonverbal behavior*, *33*(3), 149-180.

Hareli, S., Elkabetz, S., & Hess, U. (2019). Drawing inferences from emotion expressions: The role of situative informativeness and context. *Emotion*, *19*(2), 200.

Hoemann, K., Gendron, M., & Barrett, L. F. (2017). Mixed emotions in the predictive brain. *Current Opinion in Behavioral Sciences*, *15*, 51-57.

Le Mau, T., Hoemann, K., Lyons, S. H., Fugate, J., Brown, E. N., Gendron, M., & Barrett, L. F. (2021). Professional actors demonstrate variability, not stereotypical expressions, when portraying emotional states in photographs. *Nature communications*, *12*(1), 1-13.

Ong, D. C., Zaki, J., & Goodman, N. D. (2019). Computational models of emotion inference in theory of mind: A review and roadmap. *Topics in cognitive science*, *11*(2), 338-357.

Ong, D. C., Zaki, J., & Goodman, N. D. (2015). Affective cognition: Exploring lay theories of emotion. *Cognition*, *143*, 141-162.

Rule, N. O., Tskhay, K. O., Freeman, J. B., & Ambady, N. (2014). On the interactive influence of facial appearance and explicit knowledge in social categorization. *European Journal of Social Psychology*, *44*(6), 529-535.

Russell, J. A., & Carroll, J. M. (1999). On the bipolarity of positive and negative affect. *Psychological bulletin*, *125*(1), 3.

Saxe, R., & Houlihan, S. D. (2017). Formalizing emotion concepts within a Bayesian model of theory of mind. *Current opinion in Psychology*, *17*, 15-21.

Schatz, H. (2016). *In Character: Actors Acting.* Bulfinch.

Schatz, H. (2013). *Caught In the Act: Actors Acting.* Glitterati Incorporated.

Shen, X., Mann, T. C., & Ferguson, M. J. (2020). Beware a dishonest face?: Updating face-based implicit impressions using diagnostic behavioral information. *Journal of Experimental Social Psychology*, *86*, 103888.

Somerville, L. H., & Whalen, P. J. (2006). Prior experience as a stimulus category confound: an example using facial expressions of emotion. *Social cognitive and affective neuroscience*, *1*(3), 271-274.

Spunt, R. P., & Adolphs, R. (2019). The neuroscience of understanding the emotions of others. *Neuroscience letters*, *693*, 44-48.

Todorov, A., & Duchaine, B. (2008). Reading trustworthiness in faces without recognizing faces. *Cognitive Neuropsychology*, *25*(3), 395-410.

Wallbott, H. G. (1988). In and out of context: Influences of facial expression and context information on emotion attributions. *British Journal of Social Psychology*, *27*(4), 357-369.

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological science*, *17*(7), 592-598.

Wu, Y., Baker, C. L., Tenenbaum, J. B., & Schulz, L. E. (2018). Rational inference of beliefs and desires from emotional expressions. *Cognitive science*, *42*(3), 850-884.

Zhou, H., Majka, E. A., & Epley, N. (2017). Inferring perspective versus getting perspective: Underestimating the value of being in another person's shoes. *Psychological science*, *28*(4), 482-493.