This paper is accepted for publication, after peer review, in Cognition & Emotion, but it is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available at http//dx.doi.org/10.1080/02699931.2025.2578697

Observer and Stimulus Factors Jointly Shape Perceptual Similarity of Static and Dynamic Facial Emotions

Laura Manno[1], Louise Ewing[1], Stephanie Rossit[1], Andrew Bayliss[1], Guomei Zhou[2], Mintao Zhao[1]

1, School of Psychology, University of East Anglia, Norwich, UK.

2, Department of Psychology, Sun Yat-sen University, Guangzhou, China

Correspondence concerning this paper should be addressed to Mintao Zhao.

School of Psychology

University of East Anglia

Norwich Research Park

Norwich NR4 7TJ, UK

Email: Mintao.Zhao@uea.ac.uk

Abstract

Facial emotion perception is influenced by many factors (e.g., context, conceptual knowledge, facial motion), but how these factors jointly shape emotion perception remains unclear. The present study investigated how observer- and stimulus-based information interact to influence the perceived similarity of facial emotions. Participants categorized the emotion (Experiment 1) or assessed the intensity (Experiment 2) of two static or dynamic facial expressions before rating their similarity. The results showed that perceived similarity between facial expressions could be predicted by representational distances computed using stimulus properties, recognized emotions, or perceived intensity. Combining all three factors produced the best prediction performance, though the contribution of perceived intensity was more consistent for dynamic than static facial emotions. Moreover, the perception of static and dynamic facial emotions showed remarkable similarities and critical differences. All measures of similarity between facial expressions were highly correlated across dynamic and static conditions. However, dynamic facial emotions consistently elicited better categorization performance and lower perceived intensity compared to static facial emotions. These results demonstrate novel differences in the processing of static and dynamic facial emotions and provide new insights into how perceived emotion and intensity join with stimulus properties to shape our perception of facial emotion.

Keywords: emotion, facial expressions, emotion intensity, emotion categorization, similarity

Introduction

Facial expressions of emotions constitute powerful and complex non-verbal social signals that people use to read others' emotions (Barrett et al., 2019; Ekman, 1992; Jack & Schyns, 2015) and draw inferences about their goals and intentions (Buck & VanLear, 2002; Keltner & Haidt, 1999; Todorov et al., 2015; Van Kleef, 2009). Previous research suggests that perception of facial expressions is influenced by many different factors (e.g., context, culture, conceptual knowledge, facial texture and shape information, Elfenbein & Ambady, 2002; Elfenbein et al., 2007; Jack et al., 2012; Sormaz et al. 2016; Brooks & Freeman, 2018). Yet, how these factors jointly affect the perception of facial emotions has remained largely unexplored. For instance, happiness and pain can share overlapping facial configurations but convey distinct emotional content, whereas physically very distinct expressions such as crying and smiling can both reflect the same content: happiness. Moreover, although facial expressions are often dynamic and nuanced (Du et al., 2014; Delannoy & McDonald, 2009; Namba et al., 2017; Valstar et al., 2007), research on facial emotion perception remains largely relying on categorizing or matching static facial expressions according to separate and discrete emotion dimensions (e.g., happiness or fear). To understand how people perceive subtle differences and similarities of facial emotions, in the present study, we tested how stimulus (i.e., physical similarity) and observer-based information (i.e., perceived emotion and intensity) jointly contribute to the perception of facial emotions, for both static and dynamic facial expressions.

Theories of emotion perception often aim to reduce emotional experience to its elementary building blocks, such as basic emotion categories or shared underlying dimensions. For instance, theories based on emotion categories, such as the basic emotion theory (Ekman, 1992; Ekman & Cordaro, 2011; Levenson, 2011; Panksepp & Watt, 2011; Tracy & Randles, 2011), reduce the full range of expressions of emotion to a limited set of core and conceptually discrete categories (but see Barrett, 2006). These theories remain highly influential as they align well with significant emotional experiences and are linked to facial movements and neural processes that are potentially shaped by evolution and adaption (Ekman & Cordaro, 2011; Pessoa & Adolphs, 2010; Susskind et al., 2008). By contrast, theories based on emotional dimensions attribute emotion perception along a few continuous dimensions, such as valence and arousal (e.g., the affective circumplex, Russell, 1980, 2003; see also Mauss & Robinson, 2009). Discrete emotions, like fear and happiness, can then be constructed with distinct combinations of these underlying dimensions (Lindquist & Barrett, 2008).

Recent studies suggest that emotion recognition based on these basic emotions and emotional dimensions may overlook the similarities and nuanced differences between different facial expressions. Cowen and Keltner (2020; 2021) assessed the explanatory power of classical theoretical models using large datasets of facial, vocal, and video stimuli and large samples of participants. They found that the six basic emotion categories and the valence-arousal model each accounted for only about 30% of the variance in participants' emotion perception. They propose that a high-dimensional model encompassing more distinct emotions, such as pride, awe, and sympathy, is needed to effectively capture the complexity of emotional perception. Moreover, the representational boundaries between different emotional categories appear smooth and blended rather than discrete, reflecting the nuanced and overlapping nature of emotional experiences (see also Cowen & Keltner, 2017). To capture such nuanced differences and similarities in emotion perception, in the present study, we asked participants to directly evaluate the similarity of facial expressions in addition to emotion categorization and arousal rating.

Theories of emotion perception also face challenges in fully incorporating the role of facial motion in emotion perception. Previous research has shown significant morphological and dynamic differences between spontaneous and posed facial expressions (Delannoy & McDonald, 2009; Krumhuber et al., 2021; Namba et al., 2017; Valstar et al., 2007). Moreover, varying the temporal dynamics of facial movements also changes the recognition of facial emotion, in terms of both behavioural and neural responses (Bould, Morris, & Wink, 2008; Furl et al., 2020; Pollick et al., 2003). The same facial expression can convey subtly different emotional meanings (e.g., interest, pride, pleasure, and joy) by varying the duration or frequency of facial movements (Mortillaro et al. 2011). These findings suggest that facial motion constitutes a crucial signal for perception of facial emotion, particularly when information from static faces is inadequate (for reviews, see Krumhuber et al., 2013, 2023). In the present study, to further elucidate the differences and similarities between perception of static and dynamic facial expressions, we tested emotion perception using both still images and moving faces.

Perceptual similarity between static facial expressions is influenced by both stimulus- and observer-based factors. For instance, Sormaz et al. (2016) showed that facial emotion perception is influenced by stimulus-based physical cues, such as facial shape and texture information. In their study, they computed shape-related similarity between pairs of emotional faces with Procrustes analysis, and surface texture similarity using image-based correlations between pixel intensities. They found that these physical properties of stimuli strongly predicted subjective ratings of similarity between facial emotions. Murray et al. (2021) investigated how shape, surface, and conceptual information affect facial expression processing using both a perceptual task (i.e., ‘odd one out’) and an emotion categorization task. Using representational similarity analysis, they compared the performance of three models of emotion similarity. Two models were based on visual features of the stimuli: one focusing on facial shape and the other on surface texture. The third model was based on the similarity of emotion concepts inferred from the rating of animated emotional situations on six emotion dimensions (e.g., anger, disgust, Skerry & Saxe, 2015). They found that facial shape was associated with performance on the perceptual task, surface textures were linked to responses to the categorical task, and emotion concepts were related to performance on both tasks.

Observers' knowledge about emotional concepts also affects perceptual similarity of facial emotions. For example, when participants rate two emotions as conceptually similar, faces depicting these two emotions tend to be perceived as being of a corresponding level of similarity, even after controlling for stimulus-based physical similarity computed using the configuration of face movements (i.e., facial action units; Brooks & Freeman, 2018). Similarly, Skerry and Saxe (2015) found that perceived emotion (e.g., from facial expressions) and inferred emotion (e.g., based on emotion concepts or scenarios) shared partially overlapping neural codes (see also Brooks et al., 2019). These findings highlight the roles of observer-based factors, such as conceptual knowledge associated with emotional facial expressions, in perception of facial expressions of emotion.

How these different types of observer- and stimulus-based information jointly contribute to facial emotion perception remains to be elucidated. Firstly, the research mentioned above is primarily focused on perceptual similarity between static and stereotypical facial emotions. How these two types of information shape perception of dynamic facial emotions remain unclear. Secondly, previous studies have mainly concentrated on the similarity between core basic emotion categories (e.g., happiness vs fear; Sormaz et al., 2016), whether the findings apply to fine-grained within-category facial emotions (e.g., different expressions of happiness) has received little research attention. In addition, despite being one of the fundamental dimensions of emotion perception (Russell, 1980, 2003), how arousal/intensity of emotion contributes to perceptual similarity of facial emotions remains unclear. Finally, while previous research highlighted the contribution of various types of information (e.g., facial shape or conceptual knowledge about emotion) to facial emotion perception, they have rarely tested how different types of information make shared and unique contributions to emotion perception (cf. Brooks & Freeman, 2018; Murray et al., 2021). With the present study, we aimed to address these questions.

We conducted two experiments to investigate how emotional concepts, emotional intensity, and stimulus-based physical information jointly influence the perception of similarity between facial emotions, both within- and across-emotional categories, depicted in both static and dynamic facial expressions. Experiment 1 focused on emotional concepts and stimulus-based properties, and Experiment 2 focused on emotional intensity and stimuli properties. For each experiment, we first tested the association between directly rated similarity between facial emotions and the corresponding similarity computed from (a) participants' perception of emotion concepts or intensity, and (b) physical properties of images or videos showing facial emotions. We then used both representational similarity analysis and multiple linear regression analysis to examine how perceived and stimulus-based information predict directly rated similarity between facial emotions. Finally, we performed model comparison analysis to contrast regression models with different combinations of observer- and stimulus-based information, which allowed us to examine how each type of information makes unique and shared contributions to perceived similarity between facial expressions of emotions.

Experiment 1

Experiment 1 investigated how perceived emotion categories (i.e., conceptual information) and stimulus-based similarity (physical information) jointly affect the perception of similarity between facial expressions, and whether their contributions differ between perception of static and dynamic facial emotions. For the former, we specifically tested whether combining both physical and conceptual information provides a better account of perceptual similarity than when considering either alone. For the latter, we tested whether similarity between *dynamic* facial expressions is influenced by physical or conceptual information differently than *static* facial expressions. To address these questions, we asked participants to perform a combined emotion categorization and similarity rating task. Specifically, they were shown pairs of facial expressions of emotion (both static or both dynamic) sequentially and asked to categorize each facial expression in a forced choice task. Then, they rated the similarity of the pair of facial expressions. We computed *categorization similarity* across facial expressions using participants' confusions errors in the emotion categorization task (e.g., Murray et al., 2021; Skerry & Saxe, 2015) and calculated *physical similarity* across the expressions using image-based Gabor similarity (e.g., Bülthoff & Zhao, 2021; Dobs et al., 2014). If both physical and conceptual information contribute to judgments of similarity, then both categorization and physical similarity should be correlated with perceived similarity in the rating task. If these two types of information make both unique and shared contributions to perceptual similarity between facial expressions, then the combination of physical and conceptual information should yield better prediction for participants' ratings than when based on either alone.

Methods

Participants

Eighty undergraduate students from the University of East Anglia took part in the study. Data from five participants were excluded from the analysis due to their low response rates (i.e., responded to < 50% of the trials) or a request to "not use their data" during debrief. For the final sample of 74 participants, 39 completed the dynamic condition (18-33 years old, Mean age = 20.6, SD = 3.06; 5 males, 34 females; 36 European; 3 Asian), and 35 completed the static condition (18-28 years old, Mean age = 19.9, SD = 2.41; 4 males, 31 females; 33 European, 1 Asian, and 1 African). Power analysis using G*Power3.1 showed that, for a 2 by 3 mixed design, to detect a medium effect (f = 0.25) of within factor or within-between interaction at α =.05 (with assumption of correlation among repeated measures = 0.5), 32 participants in total is required to achieve a power (1-β) of .85. Here and for Experiment 2, all participants provided informed consent before taking part in the study, were debriefed at the end, and received course credits as compensation for their time. The procedure was approved by the University Ethics Committee.

Stimuli and Tasks

Our stimuli were taken from the MPI Facial Expression Database (Kaulard et al., 2012), which features non-professional actors posing facial expressions based on the description of everyday scenario prompts. To record facial expressions for the database, Kaulard et al. (2012) first told participants an scenario (e.g., “While doing sports you suddenly have an accident in which you twist one ankle and graze your knee.”) and then asked them “to remember a similar situation in their life, to imagine that they were in that situation again, and to act accordingly” in front of an array of cameras (p. 5). Participants could repeat the expression until they felt comfortable before recording. Therefore, the recorded facial expressions were posed in response to imaginary situations, which may differ from genuine emotional expressions in real life and may differ in terms of posted facial movements across actors (i.e., individual differences in the production of facial emotions).

We selected three emotions (i.e., happiness, fear, and pain) that include both positive and negative valence and both heavily and less investigated emotions that can be reliably characterized by a distinctive pattern of facial expression (Cowen & Keltner, 2020; Williams, 2002). For each emotion, we included high- and low-intensity expressions induced by two different scenarios, respectively (See **supplementary Table S1** for the list of scenarios used to induce the selected facial expressions used in the present study).

For dynamic facial expressions, we initially selected 60 video clips from 10 actors in the database (5 females and 5 males, each had 6 videos of 3 emotions × 2 intensities). We edited the videos to remove non-face related information (e.g., tracking markers in the original videos) and converted them into grayscale. All videos started with 5 frames of a neutral expression, with the face centered on a black background measuring 720 x 576 pixels. All six facial emotions from the same actor were of the same length (1000 or 2000 ms), ending at the frame showing the peak of the emotion.

Static stimuli were created by extracting the frame conveying the peak of the emotion from each video. These peaks were established from a pilot study in which ten adult volunteers categorized the emotion in each video and identified the frame they believed to convey the emotional peak. This pilot also prompted our exclusion of stimuli from one actor, because participants indicated they had difficulties accurately identifying their intended emotions. The final stimuli included a total of 54 static stimuli and 54 dynamic stimuli (9 actors × 3 emotions × 2 intensities).

Like many facial expression databases (see Krumhumber et al., 2021, for a review), not all actors in the MPI facial expression database (Kaulard et al., 2012) produced identical facial movements, even when responding to the same emotion scenarios. To evaluate potential variations across actors in posing the same facial expressions, we performed a stimulus-based analysis that infers the consistency of emotion production by examining the consistency of emotion perception using a pretrained computational algorithm (See **Supplementary Figure S1**). That is, the more similar the patterns of algorithm-based emotion perception, the more consistent the emotion production is across actors. The results showed that different actors posed the same facial expressions in a highly consistent manner (mean cosine similarity, 0.80, SD, 0.07). Moreover, this *across-actor consistency* in posing the same facial expressions was higher than the *within-actor consistency* in posing different facial emotions (mean cosine similarity, 0.59, SD, 0.10, $t(13) = 4.51$, $p < .001$). These results suggest that the production of these facial expressions is reasonably consistent across actors.

Procedure

Participants completed the integrated emotion categorization and similarity rating task (**Figure 1**) online using the Gorilla platform (Cauldron Science Ltd., UK). Each trial started with a fixation cross (1000 ms) followed by a text screen indicating whether they would see images or videos of facial expressions (e.g., “Image 1” or “Video 1”, 500 ms). The stimulus was then displayed centrally (static condition: 2000 ms; dynamic condition: 1000 or 2000 ms depending on the actors) before a response screen appeared (for a maximum of 5 seconds). Participants were asked to categorize facial expressions as accurately and quickly as possible, by pressing one of the three buttons corresponding to the three emotion categories. Once a response was made or the 5-second display time passed, the procedure was then repeated for the second stimulus, which showed a facial expression from the same actor displayed for the same duration. After both categorizations, participants were asked to rate the similarity between the two facial expressions using a 7-point Likert scale ranging from 1 (totally different) to 7 (exactly the same).

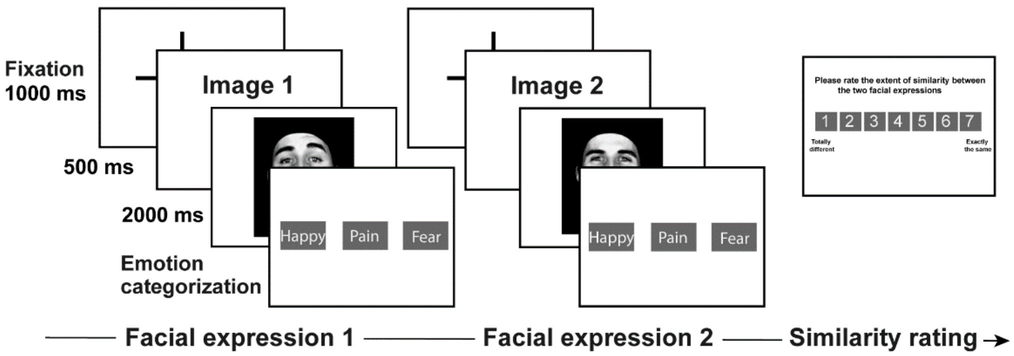

Figure 1. Example of a trial sequence used in the static condition in Experiment 1. On each trial, participants categorized two facial expressions according to their emotion and then rated the degree of similarity between the two facial expressions using a 7-point Likert scale.

Each participant completed 160 trials in a randomized order with either dynamic or static facial expressions. Emotional category and intensity of the two facial expressions (i.e., stimulus 1 and 2) varied across trials. Specifically, 40 trials showed identical facial expressions (same emotion and intensity), 40 showed the same emotion but different intensities (e.g., high intensity happy and low intensity happy), 40 showed different emotions at the same level of intensity (e.g., low intensity happy and low intensity fear), and 40 showed different emotions at different levels of intensity (e.g., high intensity happy and low intensity fear). The three facial emotions were relatively equally allocated across the 320 stimuli used for the 160 trials (Fear, 106 times; Happy, 108 times; Pain, 106 times).

Computing categorization and physical similarity between facial expressions

By examining behavioral confusions between facial expressions and recognized emotions, previous studies have used emotion categorization responses to establish the conceptual similarity between different facial emotions, (e.g., Brooks et al., 2019; Skerry & Saxe, 2014; Murray et al., 2021). Following this approach, we asked participants to perform a similar forced-choice emotion categorization task and calculated the *categorization similarity* based on their confusion errors observed in this task. We used categorization similarity to index conceptual similarity between facial expressions. That is, the more the emotions of the two facial expressions were confused with each other during emotion categorization, we inferred the more similar they are in terms of participants' concepts about these emotions. Note that emotion categorization made during this forced-choice task may not capture the full "contents" of facial emotion perception.

Physical similarity between two facial expressions was calculated as the Gabor similarity between two stimuli (Lades et al., 1993). Prior works have shown that Gabor similarity between facial images correlates with perceptual similarity of facial identities (Bülthoff & Zhao, 2020, 2021; Yue et al., 2012), expressions (Xu & Biederman, 2010), and facial movements (Dobs et al., 2014). Gabor similarity has also been used to provide an objective similarity measure of facial emotion stimuli (images: Susskind et al., 2007; videos, Dobs et al., 2014).

For the static condition, we first autodetected faces using a face-detector embedded in the Matlab computer vision toolbox (MathWorks, MA) and resized autodetected face images into 256 by 256 pixels. We then generated a feature vector for each image with 8,000 filters, which resulted from applying a Gabor jet [5 scales × 8 orientations × 2 phases (sine and cosine) at the intersections of a 10 × 10 uniform grid. *Gabor dissimilarity* was computed as the Euclidean distance between the two feature vectors generated for each of the two facial images. Identical images produced a value of zero, and more different images produced higher Gabor dissimilarity values. For the dynamic condition, we calculated Gabor dissimilarity between corresponding individual frames of the two videos, using the average values across all frames as the measure of physical similarity between two videos (e.g., Dobs et al., 2014).

Results and Discussion

Emotion categorization

Error rates for emotion categorization are shown in **Figure 2A.** A mixed 2 (condition: dynamic vs. static; between-participants factor) by 3 (emotion: happy, fear, pain; within-participants factor) ANOVA revealed a main effect of facial emotion, $F(2, 144) = 121.02$, $p < .001$, $\eta_p^2 = .63$. Follow up contrasts showed that participants made fewest errors when categorizing facial expressions of happy (M±SE: 0.08±0.01), followed by fear (0.17±0.01), and then pain (0.34±0.02, for both the static condition, all $t(34) > 5.39$, $p < .001$, *Cohen's d* > 0.91, and the dynamic condition, all $t(38) > 6.26$, $p < .001$, *Cohen's d* > 1.00. The main effect of stimulus condition was significant, $F(1, 72) = 10.20$, $p = .002$, $\eta_p^2 = .12$. Participants made fewer categorization errors in the dynamic condition (0.17±0.01) than those observed in the static condition (0.22±0.01). The interaction between facial emotion and stimulus condition was not significant, $F(2, 144) = 0.58$, $p = .561$, $\eta_p^2 < .01$, suggesting that the effect of emotion was consistent across static and dynamic facial expressions. These results demonstrate a dynamic advantage in facial emotion recognition and that facial expressions of happiness are more separable from those of fear and pain.

To explore the influence of individual differences in emotion recognition (i.e., variations across participants) and emotion production (i.e., variations across the actors posing the facial expressions) on our results, we conducted a linear mixed model (LMM) analysis, with both participants and actors of stimuli as random factors ('Response ~ Condition * Emotion + 1|Subject

+ 1|StimuliActor’). The results showed significant effects of stimulus condition, $F(1, 1994) = 6.82$, $p = .009$, facial emotion, $F(1, 1994) = 30.87$, $p < .001$, and a non-significant interaction, $F(1, 1994) = 1.70$, $p = .192$. Therefore, our results cannot be attributed to individual differences in emotion categorization among our participants, nor to variations in emotion production across the actors posing these facial expressions.

When examining confusion errors, where participants mistakenly categorized target facial emotion as other emotions (**Figure 2b**), we found that fear was confused more often with pain than happy (dynamic condition: $t(38) = 6.68$, $p < .001$, *Cohen’s d* = 1.07; static condition, $t(34) = 2.79$, $p = .009$, *Cohen’s d* = 0.46). Similarly, pain expression was confused more often with fear than happy (dynamic condition: $t(38) = 8.64$, $p < .001$, *Cohen’s d* = 1.38; static condition: $t(34) = 5.15$, $p < .001$, *Cohen’s d* = 0.87). Happy was confused more often with pain than fear [dynamic condition, $t(38) = 4.69$, $p < .001$, *Cohen’s d* = 0.75; static condition, $t(34) = 3.90$, $p < .001$, *Cohen’s d* = 0.66]. Moreover, the proportion of confusion errors tended to be more evenly distributed in the static condition than in the dynamic condition (fear, $\chi^2(1) = 6.43$, $p = .011$; happy, $\chi^2(1) = 2.54$, $p = .111$; pain, $\chi^2(1) = 1.59$, $p = .208$). These results indicate that facial expressions of fear and pain are more frequently confused with each other, and less so with facial expressions of happiness.

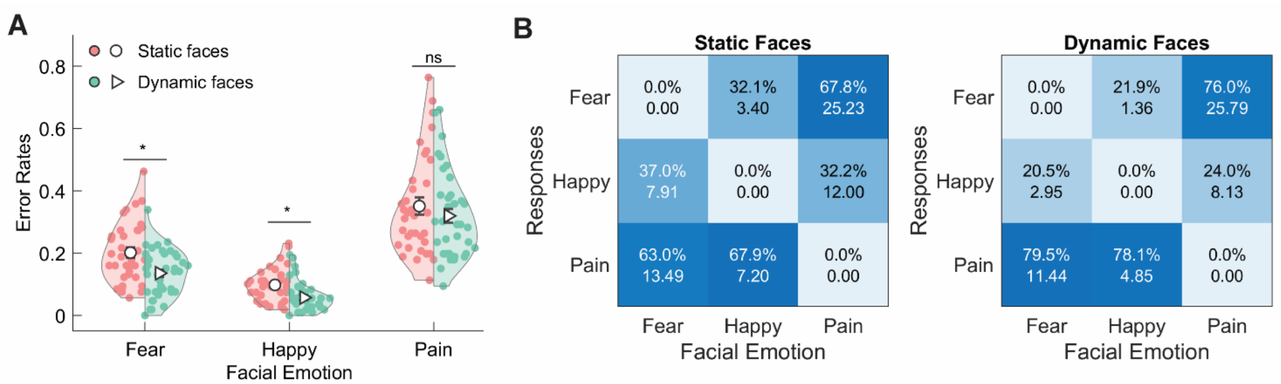

Figure 2. Emotion categorization performance in Experiment 1. A. Mean error rates for categorizing static and dynamic facial emotions. Dots represent individual participants’ performance. *, $p < .001$. B. Proportion and mean number of confusion errors made during emotion categorization. X-axis represents facial emotion, and Y-axis represents categorized emotion by participants.

Association between perceptual similarity and emotion categorization

To examine how the recognized emotions from the two facial expressions on each trial contributed to their rated similarity, we computed the categorization similarity for each face pair in the 160 trials. Specifically, for the two facial expressions in a trial, we calculated how likely each of these two faces was wrongly categorized as showing the intended emotion of the other face and then averaged the two values. The resulting value indexes how likely the two facial expressions were confused with each other in terms of emotion categorization. We then correlated the mean categorization similarity observed for each of the 160 trials with the corresponding mean perceptual similarity derived from the similarity rating task (**Figure 3A**). We found strong correlations between categorization and perceptual similarity, for both the static (r =.81, p < .001) and dynamic conditions (r = .86, p < .001). Moreover, this association was consistently observed for individual participants, whether they viewed static (all r > .42, p < .001, except one participant, r = .08, p= .293) or dynamic facial expressions (all r > .52, p < .001, **Figure 3B**). The more the two faces induced confusion in their emotion categorization, the higher their rated similarity.

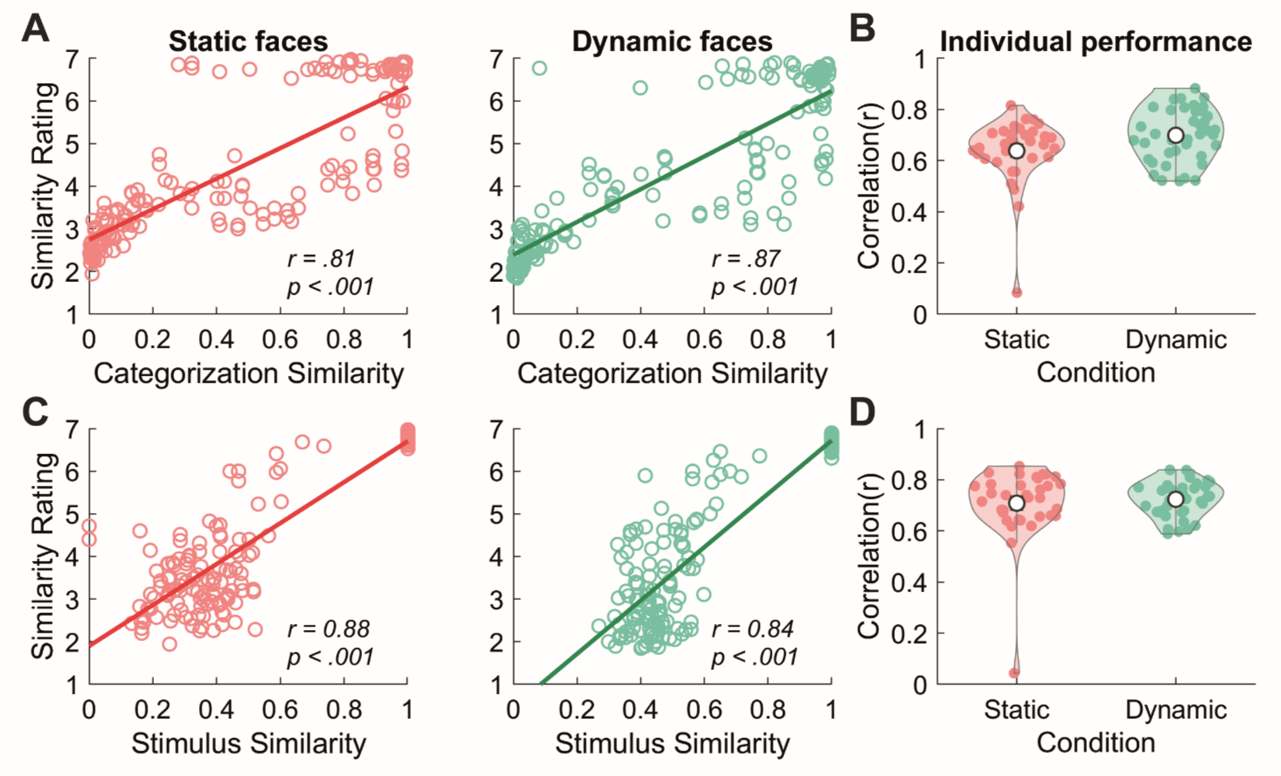

Figure 3. Perceptual similarity between facial expressions is associated with emotion categorization similarity and stimulus-based physical similarity, at both group level (A, C) and individual participants level (B, D). Dots in A and C represent measures observed for individual trials used in the Similarity Rating task; dots in panels B and D represent results of individual participants.

Association between perceptual similarity and stimuli properties

Similar analyses were performed to investigate how physical similarity between two facial expressions contributed to their rated similarity. To do so, we first computed Gabor dissimilarity between two images or videos in a trial (see Methods). We then standardized the measure [i.e., physical similarity = (Maximal observed Gabor dissimilarity – Gabor dissimilarity) / Maximal observed Gabor dissimilarity], so the resulting *physical similarity* values would vary between 0 and 1, with 1 representing the highest stimulus similarity (i.e., identical stimuli). As shown in **Figure 3C**, we observed significant correlations between stimulus-based physical similarity of the 160 trials and its corresponding perceptual similarity derived from rating, for both the static (r = .88, p < .001) and dynamic conditions (r = .84, p < .001). Analysis of data from individual participants confirmed that this correlation occurred for all but one participant (static condition, all r > .55, p <.001, except one participant, r = .04, p =.602; dynamic condition, all r > .58, p <.001, **Figure 3D**).

Representational similarity metrices across facial expressions of emotion

To examine the relative similarity across all six facial expressions (3 emotions × 2 intensities), we computed their representational similarity based on perceptual similarity, categorization similarity, and physical similarity. Across the 160 trials, we averaged these measures according to the full combination of the six facial expressions, yielding a 6 by 6 matrix for each of the three measures (**Figure 4**; see supplementary **Table S2** for full details of correlation analysis). Representational similarity matrices for static and dynamic faces were remarkably similar for all three measures (all r > .97, p <.001). Categorization and perceptual similarity matrices appeared to be more aligned with each other for dynamic faces (r =.93, p <.001) than for static faces (r =.88, p <.001). A reversed pattern was observed for physical and perceptual similarity, with numerically higher correlation for static faces (r =.94, p <.001) than for dynamic faces (r =.90, p <.001). Both perceptual and categorization similarity matrices showed distinctive grouping of facial emotions (regardless of intensity), which was not clearly visible in the physical similarity matrices.

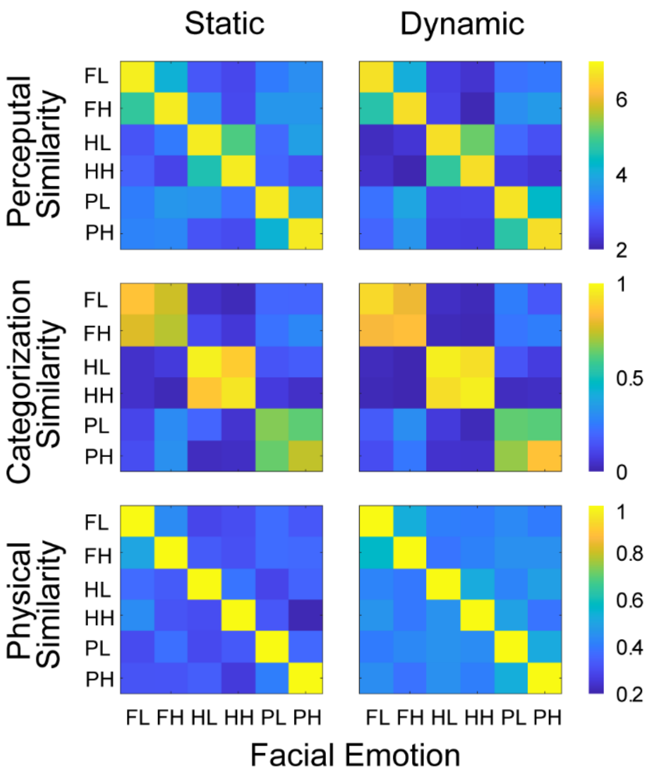

Figure 4. Representational similarity matrices for static and dynamic facial expressions based on perceptual, categorization, and physical similarity. Facial emotions from top to bottom (Y axis) and from left to right (X axis) represent low (L)- and high (H)-intensity expressions of fear (FL, FH), happiness (HL, HH), and pain (PL, PH). Blue to yellow color bars represent low to high similarity.

Shared and unique contribution of categorization and physical similarity to perceptual similarity

Multiple linear regression analyses were performed to examine how physical and categorical similarities predict the rated similarity for the 160 trials. As summarized in **Table 1**, both categorization and physical similarity predicted perceptual similarity well (all adjusted $R^2 \geq .66$, all $F \geq 304.52$, $p < .001$). Physical similarity accounted for a greater proportion of the variance than categorization similarity for static faces (adjusted R^2, .77 vs .66), but both predictors accounted for similar proportions of the variances for dynamic faces (adjusted R^2, .76 vs .75). Importantly, combining both predictors yielded the best model performance, accounting for 90% and 91% of total variances observed for the rating of static and dynamic faces respectively. Model comparisons revealed better performance when both physical and categorization similarity were used as predictors than when using either alone (all R^2 changes > 0.12, all F changes > 184.02, p <.001). These results indicate that both categorical and physical information contribute to perceived similarity between facial expressions, and each makes unique and shared contributions.

Table 1. Results of regression analysis for Experiment 1

Model	R^2	$R^2_{Adjusted}$	*F*	*ΔF-1*	*ΔF-2*
		Static facial emotion			
1. Categorization similarity	.658	.656	304.52*		
2. Physical similarity	.775	.774	545.07*		
3. Combined	.897	.895	680.23*	361.38*	184.02*
		Dynamic facial emotion			
1. Categorization similarity	.750	.749	474.70*		
2. Physical similarity	.765	.764	515.10*		
3. Combined	.913	.912	823.63*	293.57*	266.52*

*Note. *, p <.001; ΔF–1/2, change of F values when compared to Model 1 or 2.*

Experiment 2

Experiment 1 showed that both physical similarity and emotion categorization similarity contribute to how we perceive facial emotion, and that facial motion provides an advantage in recognizing emotions. In Experiment 2, we investigated how emotion intensity information affects perceptual similarity between facial emotions. This question is critical because emotion intensity is related to one of the fundamental dimensions of the valence-arousal account of emotion perception (Mauss & Robinson, 2009; Kuppens et al., 2013; Russel, 2003). Furthermore, processing of emotion intensity has been shown to involve dissociable neural substrates from those supporting emotion categorization (e.g., Muukkonen & Salmela, 2022). To this end, we asked participants to rate the intensity of facial expressions before judging their similarity.

We manipulated emotion intensity differently from previous studies, which have typically artificially varied facial emotion stimuli. For instance, different levels of emotion intensity are often achieved by exaggerating or weakening facial muscle movements via morphing between images showing different facial emotions (e.g. Biele & Grabowska 2006; Calvo et al., 2016; Murray et al., 2021; Sormaz et al., 2016). Although these manipulations provide elegant control over stimulus-based properties, they may not capture emotional intensity expressed spontaneously in facial expressions (e.g., Chen et al., 2024). In this experiment, the intensity of facial emotions was defined by the scenarios the actors used for posing specific facial emotions (Kaulard et al., 2012).

Emotion intensity conveyed in these posed facial expressions is arguably less artificial than that via directly manipulating facial images and videos.

For the similarity rating task, we changed the 7-point discrete Likert scale to a continuous 0-100 scale. This allowed us to examine whether perceptual similarity obtained in Experiment 1 persists when participants make more nuanced comparisons between facial expressions. Finally, by integrating results from Experiment 1, Experiment 2 also allowed us to further examine how emotion categorization, intensity perception, and physical similarity jointly contribute to perceptual similarity between facial expressions of emotions. If perceived intensity information makes both unique and shared contributions to perceptual similarity between facial expressions, we should observe a strong association between intensity similarity and rated similarity and significantly improved model prediction performance when intensity information is included as additional predictors.

Methods

Participants

Forty students from the University of East Anglia were recruited for this experiment (11 males, 29 females; aged between 19-30 years, M = 20.8, SD = 2.97; 37 European, 2 Asian, and 1 American; 20 each were randomly assigned in the static and dynamic conditions). Using the same data quality criteria from Experiment 1, no participants were excluded from the data analysis. Power analysis using G*Power3.1 showed that, for a 2 by 2 mixed design, to detect a medium effect (f = 0.25) of within factor or within-between interaction at $\alpha = .05$ (with assumption of correlation among repeated measures = 0.5), 38 participants in total is required to achieve a power ($1-\beta$) of .85.

Stimuli, Tasks, and Procedure

The stimuli were the same as in Experiment 1. The tasks and procedure were also similar to those of Experiment 1, with two variations (**Figure 5**). First, for each of the two facial expressions in a trial, instead of performing the emotion categorization task, participants were asked to rate the intensity of each facial expression using a 7-point Likert scale. Participants made their response by clicking on the corresponding buttons on the screen. For the similarity rating task, participants made their judgments by moving the handle of an on-screen slider, where 0 represents “Totally different” and 100 “Exactly the same”. The handle was initially placed on 50.

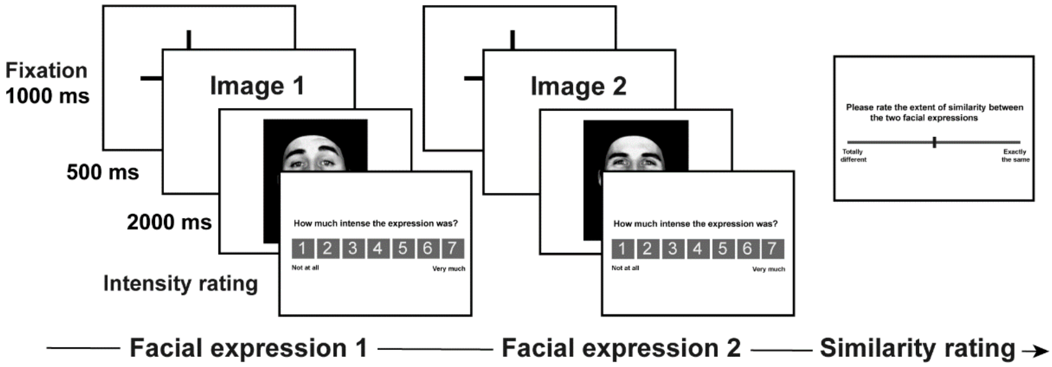

Figure 5. Example of the emotion intensity and similarity rating task for the static condition in Experiment 2. In each trial, participants first rated the intensity of two facial expressions before rating their similarity using a slider ranging from 0 (Totally different) to 100 (Exactly the same).

Results and Discussion

Perception of emotional intensity

Mean rated intensities for dynamic and static facial expressions are shown in **Figure 6**. A mixed 2 (condition: static vs dynamic; between-participants factor) × 2 (stimulus intensity: low vs high; within-participants factor) ANOVA revealed a significant main effect of stimulus intensity, $F(1,38) = 139.94$, $p < .001$, $\eta_p^2 = .79$. This result confirmed that facial expressions elicited by high intensity emotional scenarios (M±SE = 4.17±0.11) were perceived as more intense than those elicited by low intensity emotional scenarios (3.90±0.11). The main effect of condition was also significant, $F(1,38) = 5.54$, $p = .024$, $\eta_p^2 = .13$, with static facial expressions (4.30±0.16) perceived as more intense than dynamic facial expressions (3.77±0.16). There was no significant interaction between stimulus intensity and condition, $F(1,38) = .070$, $p = .793$, $\eta_p^2 = .002$. As in Experiment 1, to examine whether our findings persist after accounting for both individual differences in emotion perception across participants and in emotion production across the actors posing the facial expressions, we performed a LMM analysis, with participants and actors of the stimuli as random factors (i.e., 'Response ~ Condition * Intensity + 1|Subjects + 1|StimuliActors'). The results revealed the same pattern of responses (stimulus intensity, $F(1,716) = 5.36$, $p = .021$; condition, $F(1,716) = 4.63$, $p = .032$; Interaction, $F < 1$). Therefore, the effects of facial motion and scenario intensity on emotion perception are unlikely driven by individual differences in emotion categorization or actor-related variability in the production of these facial expressions.

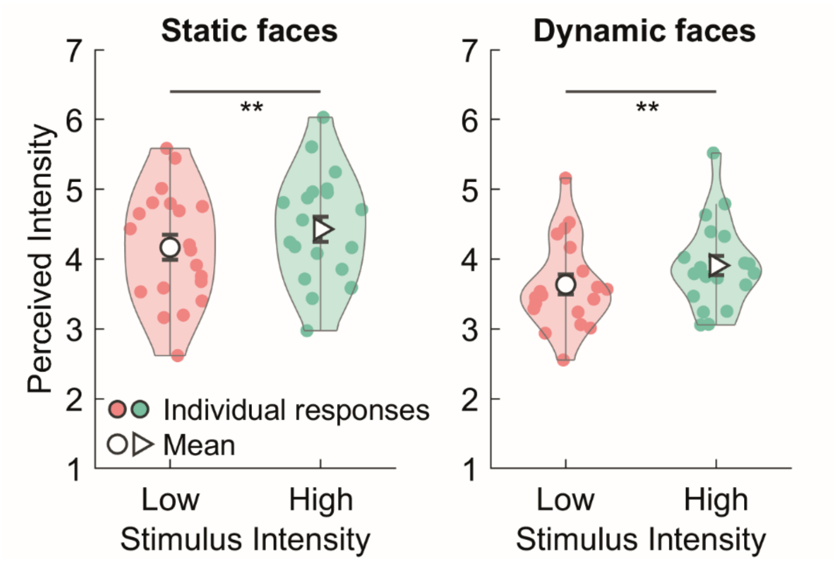

Figure 6. Perceived emotional intensity from static and dynamic facial expressions elicited by low- and high-intensity emotional scenarios. **, *p* < .001.

Association between perceived intensity, stimuli properties and rated similarity

To examine whether the rated similarity between two facial expressions is associated with the perceived intensity from each facial expression, we first computed the difference between the rated intensity for the pairs of facial expressions in each trial. We then standardized these differences [i.e., Intensity similarity = (Maximal observed intensity difference – intensity differences) / Maximal observed intensity difference]. The resulting *intensity similarity* measures would vary between 0 and 1, with 1 being the highest intensity similarity (i.e., identical rating of intensity). We found that mean intensity similarity was significantly correlated with mean rated similarity across the 160 trials, for both the static (r = .83, p < .001) and dynamic conditions (r = .81, p < .001, **Figure 7A**). Significant correlation between intensity similarity and perceptual similarity was also observed for each participant in both the static and dynamic conditions (all r > .27, p < .002; **Figure 7B**).

Consistent with Experient 1, we observed strong correlations between the physical similarity computed for each pair of facial expressions and its corresponding perceptual similarity obtained from ratings, for both the static (r = .89, p < .001) and dynamic conditions (r = .88, p < .001, **Figure 7C**). This correlation was also observed for individual participants (all r > .38, p < .001; **Figure 7D**).

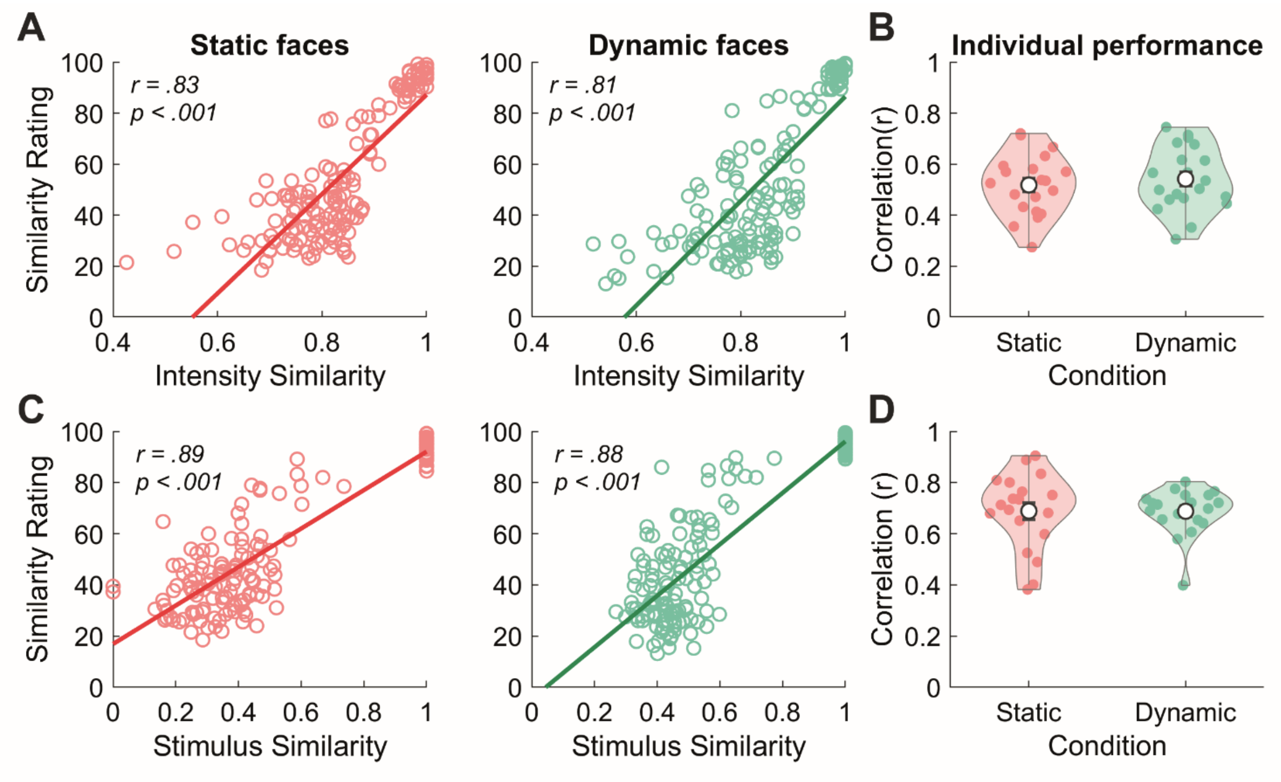

Figure 7. Rated similarity is associated with intensity similarity (A, B) and stimuli-based physical similarity (C, D). Dots in A and C represent measures obtained from individual trials. Dots in C and D represent correlation coefficients observed from individual participants.

We also computed representational similarity metrices across the six facial expressions of emotions to examine their relative similarity (**Figure 8**, see **Table S3** for full details of correlation analysis). Static and dynamic facial expressions showed remarkably similar matrices for all three measures (all r > .86, p <.001). Perceptual similarity matrices appeared to be more aligned with physical similarity matrices (static faces, r =.95, dynamic faces, r =.91; both p <.001) than with intensity similarity matrices (static faces, r =.91, dynamic faces, r =.88, both p <.001). Note that only perceptual similarity matrices showed distinctive grouping of facial emotions (regardless of intensity), which was not clearly visible in the intensity and physical similarity matrices.

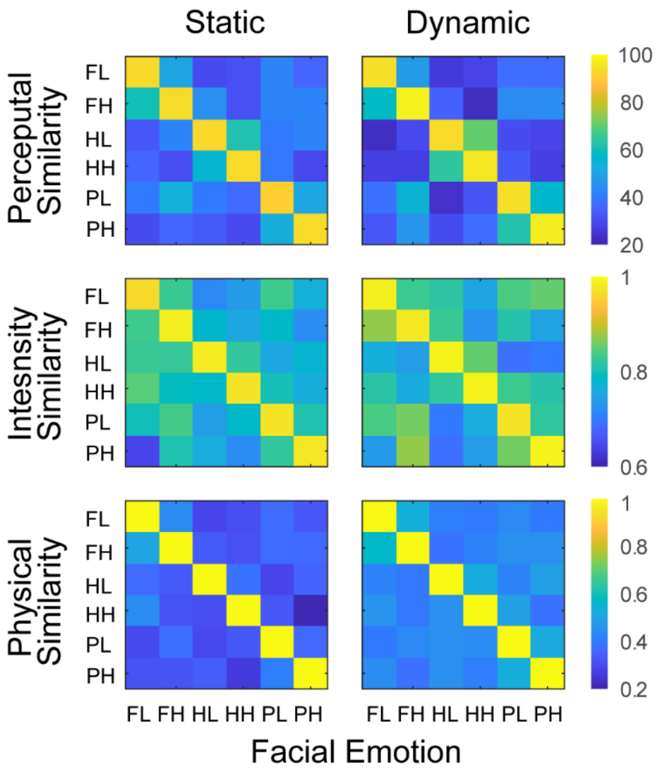

Figure 8. Representational similarity matrices for static and dynamic facial expressions based on perceptual, intensity, and physical similarities. Facial emotions from top to bottom (Y axis) and from left to right (X axis) represent low (L) and high (H) intensity expressions of fear (FL, FH), happiness (HL, HH), and pain (PL, PH). Blue to yellow color bars code low to high similarity values.

Shared and unique contribution of intensity and physical similarity to perceptual similarity

As in Experiment 1, multiple linear regression analyses were performed to examine how physical and intensity similarity predict similarity judgements. As summarized in **Table 2**, both intensity and physical similarity predicted perceptual similarity well (all adjusted $R^2 \geq .68$, $F \geq 294.85$, $p < .001$). Physical similarity accounted for more proportion of the variance than intensity similarity for static faces (adjusted R^2, .80 vs .68), but they accounted for similar proportion of variances for dynamic faces (adjusted R^2, .77 vs .75). The combined model incorporating both physical and intensity information showed the best prediction performance, accounting for 83% and 82% of rating variances observed in the static and dynamic condition, respectively. Model comparisons showed a small but significant improvement in performance when intensity information is added to physical information (static condition: $\Delta R^2 = .03$; dynamic condition: $\Delta R^2 = .05$; both $F > 27.73$, $p < .001$), and a relatively larger improvement in performance when physical information is added to intensity information (static condition: $\Delta R^2 = .13$; dynamic condition: $\Delta R^2 = .07$; both $F > 133.82$, $p < .001$). These findings indicate that intensity and physical information make both unique and shared contributions to the perceived similarity between facial emotions.

Table 2. Results of regression analysis for Experiment 2

Model	R^2	$R^2_{Adjusted}$	*F*	*ΔF-1*	*ΔF-2*
		Static facial emotion			
1. Intensity similarity	.681	.679	337.79*		
2. Physical similarity	.798	.796	622.50*		
3. Combined	.828	.826	377.78*	133.82*	27.73*
		Dynamic facial emotion			
1. Intensity similarity	.750	.749	294.85*		
2. Physical similarity	.770	.768	527.83*		
3. Combined	.821	.818	359.13*	148.37*	44.64*

*Note. *, p < .001; ΔF–1/2, change of F values when compared to Model 1 or 2.*

Predicting perceptual similarity using physical, categorization and intensity information

To further investigate how physical, categorization and intensity information jointly contribute to perceptual similarity between facial expressions of emotion, we combined the data from the two experiments and performed multiple linear regression analysis to predict rating responses using the three predictors. As shown in **Table 3**, adding categorization similarity to physical similarity significantly improved prediction performance, for ratings observed for both the static and dynamic conditions in Experiments 1 and 2 (all $\Delta R^2 \geq .08$, $F \geq 139.68$, p <.001). Further addition of intensity similarity consistently improved the prediction performance for ratings of dynamic expressions in both experiments (all $\Delta R^2 \geq .01$, $F \geq 30.01$, p <.001). However, to predict ratings of static expressions, further addition of intensity similarity only significantly improved performance for Experiment 2 ($\Delta R^2 = .02$, $F(1,156) = 30.36$, p <.001) and not Experiment 1 ($\Delta R^2 = .002$, $F(1,156) = 3.17$, p =.077). Therefore, while combining all three predictors yielded the highest prediction performance for both static and dynamic facial expressions (all adjusted $R^2 \geq .90$, $F(3,156) \geq 460.79$, p <.001), the unique contribution of intensity information was relatively low (accounting for 2% of variances or less). Detailed comparisons across all combination of the three predictors indicated that the effect of intensity similarity was largely accounted for by physical similarity between facial expressions (see supplementary **Table S4** for full details of model prediction performance and model comparisons across all combination of these three predictors).

Table 3. Prediction of similarity ratings using physical (P), categorization (C), and intensity (I) similarity.

Model	*Predicting Ratings of Experiment 1* R^2	R^2_{Adj}	ΔR^2	ΔF	*Predicting Ratings in Experiment 2* R^2	R^2_{Adj}	ΔR^2	ΔF
				Static condition				
P	.775	.774	.775	545.07*	.798	.796	.798	622.50*
P+C	.897	.895	.121	184.02*	.879	.878	.082	571.27*
P+C+I	.899	.897	.002	3.17^{ns}	.899	.897	.020	30.36*
				Dynamic condition				
P	.765	.764	.765	515.10*	.770	.768	.770	527.83*
P+C	.913	.912	.148	266.52*	.878	.877	.108	139.68*
P+C+I	.927	.926	.014	30.01*	.901	.899	.023	35.72*

*Note. $\Delta R^2/\Delta F$, change of R^2 and related F statistics when compared to the previous model (e.g., P vs intercept, P+C vs P). *, p < .001. ns, not significant.*

General discussion

In the present study, we investigated how emotion recognition, intensity perception, and stimulus-based physical information jointly contribute to the perception of facial expressions of emotion. Unlike previous research that focuses on how to map facial expressions onto semantically discrete emotion categories (i.e., emotion recognition and categorization), we aimed to understand how these different types of information jointly predict perceptual similarity between scenario-based facial expressions. Our results demonstrate that both emotion categorization (as posited by the Basic Emotion Theory, Ekman, 1992; Levenson, 2011) and intensity perception (as proposed by the Dimension Theory of Emotion, e.g., Russel, 1980, 2003; see also Mauss & Robinson, 2009) shape the judgment of similarity between facial expressions of emotion. Furthermore, these two types of observer-based information also interact with stimulus-based information (i.e., physical similarity) to better predict perceptual similarity judgments. While the unique effect of emotion categorization was clearly dissociated from the effect of stimulus-based physical information, the unique contribution of intensity perception to similarity judgments was less so, with most of its effect accountable by stimuli properties. These findings indicate that the perception of facial emotion is more complex and nuanced than mapping facial expressions onto discrete emotion concepts (Cowen & Keltner, 2020, 2021; Shuman et al., 2015; Snoek et al. 2023).

The present study shows remarkable similarity but also critical differences between the perception of dynamic and static facial emotions. Across the two experiments, perceptual similarity between facial expressions obtained with dynamic and static facial emotions were highly correlated, at both the condition level and the trial level (supplementary **Table S2** and **S3**). This finding suggests that dynamic facial cues do not change the relative "mental distance" between two facial emotions. On the other hand, dynamic and static facial expressions did differ in terms of recognized emotions and perceived intensity. We observed a dynamic advantage in facial emotion categorization, consistent with previous research (Krumhuber et al., 2023; cf. Fiorentini & Viviani, 2011). This dynamic advantage applies to the core basic emotions (e.g., happy and fear) and other scenario-based facial expressions like pain. Previous research suggests that the beneficial effect of dynamic cues is most prominent in suboptimal situations (e.g., when facial emotions are displayed with point-light or blurred stimuli, Krumhuber et al., 2013), and it disappears when sufficient facial information is available or when facial emotions are displayed at high intensity (Bould et al., 2008; Fiorentini & Viviani, 2011). Our results indicate that the dynamic advantage is not contingent on suboptimal conditions; it persists across both high- and low-intensity emotions even when full facial information is available. Moreover, emotion categorization errors appeared to be more clustered around one of the two incorrect categories for dynamic than for static facial expressions (**Figure 2B**). This polarized pattern suggests that dynamic facial cues provide strong, albeit sometimes misleading, cues that guide participants' interpretation of facial emotion. As we only tested a limited set of facial expressions (i.e., high- and low-intensity expressions of happiness, fear and pain), further research is needed to determine how well these findings generalise to other basic emotion categories or other types of facial expressions (e.g., conversational or compound emotions; Dobs et al., 2014; Du et al., 2014).

Our participants also rated static facial expressions as more intense than dynamic ones. This finding concurs with a recent study (Krumhumber et al., 2021) but it contrasts with prior research reporting higher intensity, arousal, and authenticity judgments for dynamic than static facial expressions (Biele & Grabowska 2006; Krumhuber et al., 2013, but see Becker et al., 2024). Two possibilities might account for the lower intensity perceived from the dynamic than from the static facial expressions. First, peak emotions are often displayed with shorter time for dynamic (i.e., in a fraction of frames) than for static facial expressions. Participants might have aggregated the intensity perceived from individual frames to derive the overall intensity of dynamic facial expressions, which in theory would be lower than that for the static facial expressions (i.e., lasting frame of peak emotions). Second, it might be due to the different temporal profiles of emotion intensity between the spontaneous facial expressions used here and the posed, prototypical expressions used in previous studies. It has been shown that the velocity and sequence of facial movements differ between spontaneous and posed expressions (Namba et al., 2017). Posed or artificially created facial emotions typically follow a clear onset-apex-offset trajectory, where facial muscle contractions accumulate to a peak before subsiding. In contrast, spontaneous expressions often exhibit more complex temporal profiles characterised by having multiple peaks or lower-intensity emotional peaks, resembling an onset-apex-onset-apex-offset progression (Delannoy & McDonald, 2009). Participants' intensity ratings might have biased by those multiple low-intensity peaks, resulting in an overall lower rating than that observed for the static facial emotions with one high-intensity peak. This account may partially explain Becker and colleagues' (2024) recent findings. They showed that video recordings and photographs of facial emotions had similar intensity ratings, and both were perceived as more intense than artificially created dynamic facial emotions (i.e., via morphing). These results suggest that dynamic facial expressions are not consistently perceived as more emotionally intense than static facial expressions, even though they enhance emotion recognition (Calvo et al, 2016).

Our study shows that discrete emotion categorization and gradient intensity perception contribute to judgments of perceptual similarity between emotional expressions, and that these observer-based processes join with stimulus-based physical properties to shape how people perceive facial emotions. Previous research also showed that perceptual similarity between facial emotions is associated with stimulus-based properties (e.g., facial shape and texture information, Murray et al., 2021; Sormaz et al., 2016; but see Brook et al, 2018, when physical similarity of faces was measured using the facial action coding system). Nonetheless, these studies only examined the perception of static faces. Dobs et al. (2014) showed that Gabor similarity calculated between dynamic facial expressions is related to their perceived similarity, but this study assessed perceptual similarity between facial expressions only indirectly (participants decided which one of two presented facial expressions was more similar to a previously viewed one). Indirect measures of perceptual similarity between static facial expressions were also used in Murray et al. (2021, with an odd-one-out task) and in Brook et al (2018, with on-screen trajectories of mouse responses). Note that the dynamic facial expressions used in these studies are often created by animating or morphing facial images (e.g., Calvo et al., 2016; Dobs et al., 2014; Murray et al., 2021). Becker et al (2024) recently showed that dynamic morphed emotions differed from spontaneous dynamic facial emotions in terms of perceived intensity and genuineness. Here, our results demonstrate that stimulus-based physical similarity predicts well the directly rated similarity of both static and dynamic facial emotions.

For both static and dynamic facial expressions, perceptual similarity judgments of facial expression pairs were significantly correlated with their similarity when calculated using the confusion errors in emotion categorization (i.e., categorization similarity), using differences between rated intensity (i.e., intensity similarity), and using image-based Gabor dissimilarity (i.e., physical similarity). Thus, perceptual similarity between facial expressions is not exclusively determined by stimulus-based properties. The way in which people interpret the emotion and intensity of facial expressions also contributes to perceived similarity, with the combination of these two factors accounting for 80-86% of variances ratings observed in our study (supplementary **Table S4**). When both observer- and stimulus-based information are taken into consideration, prediction of perceptual similarity is better than when relying on either type of information alone, accounting for a higher proportion (90-93%) of rating variances (**Table 3**). These results are consistent with previous findings that observers' knowledge of emotional concepts influences their perception of facial expressions (Brooks & Freeman, 2018; Murray et al., 2021; Nook, Lindquist, & Zaki, 2015). For instance, Brooks and Freeman (2018) showed that when participants' conceptual knowledge of one emotion category overlapped with an alternative category, presenting facial expressions of either emotion is likely to induce co-activation of both categories, causing confusion in emotion categorization (see also Skerry & Saxe, 2014). In our study, facial emotions that are frequently miscategorized as a different emotion were also rated as being more similar to that alternative. These results suggest that conceptual information about emotions affects how people perceive the similarity between facial expressions, whether such perceptual similarity is measured implicitly (e.g., via motor responses, Brooks & Freeman, 2018, visual pattern of neural responses, Skerry & Saxe, 2014) or explicitly via direct rating.

Although perceived emotion and intensity both contribute to the rated similarity between facial expressions and enhance the prediction of ratings when combined with stimulus-based information, they interact differently with stimulus-based information in predicting perceptual similarity (supplementary **Table S4**). For instance, emotion categorization similarity consistently makes a unique contribution to rated similarity; it accounted for an additional 7-15% of response variances when combined with either physical or both physical and intensity information. In contrast, intensity information can only account for an additional 0-5% of variances when combined with physical information or with both physical and emotion categorization information. These results suggest that the contribution of intensity information is largely accounted for by physical and emotional information. Moreover, while adding emotion categorization information significantly improved prediction performance for both static and dynamic conditions across the two experiments, adding intensity information failed to improve model performance when predicting rating of static faces in Experiment 1. These results suggest that physical and emotional information may be better at accounting for intensity information conveyed via facial configuration (in static condition) than that conveyed via facial motion (in dynamic condition). Consistent with this view, Chen et al. (2024) recently showed that facial movements convey information about both emotion categories and intensity of facial expressions, and these two types of information are expressed with different temporal dynamics.

Our study may also shed light on how different types of observer- and stimulus-based information jointly predict facial expression similarity. Perceptual similarity between basic emotions has been associated with stimulus-based properties (e.g., facial shape and texture) and participants' knowledge about emotions (e.g., Brooks & Freeman 2018; Murray et al., 2021; Sormaz et al., 2016). Here, we show that these findings apply to both static and dynamic facial emotions and for both across- and within-category emotion judgments. Moreover, in addition to emotional concepts, emotion intensity also contributes to perceptual similarity between facial emotions, particularly when intensity is conveyed via dynamic facial information. Our comparison of different prediction models demonstrates that emotion categorization, intensity perception, and stimulus-based properties make both unique and overlapping contributions to the perception of facial emotion similarity. By combining observer- and stimulus-based information, our regression models exhibit remarkable prediction performance (adjusted $R^2 \geq .90$), particularly when taking potential individual differences in similarity ratings into consideration (cf. $R^2 = .386$ for the perceptual task reported in Murray et al., 2021).

Note that the use of scenario-based facial expressions may impose certain limitations on the generalizability of our findings. Firstly, caution should be taken when attempting to generalize the present findings to the perception of genuine facial expressions of emotions. Posed facial expressions (e.g., imagining being in a fearful situation) often differ from real-life genuine emotional expressions (e.g., encountering a suddenly appearing brown bear). Genuine facial expressions often involve involuntary muscle movements and are typically linked to neural processes in subcortical brain areas (Ekman, 1992; Pessoa & Adolphs, 2010), which may not be accurately reproduced via consciously controlled facial movement in posed expressions (e.g., McLellan et al., 2010). Secondly, individual differences in both the *perception* and *production* of facial expressions should be considered when investigating the contribution of observer and stimulus factors to facial emotion perception. From the stimulus side, different actors may portray the same emotion with substantial variability in their facial movements, even when they are attempting to pose the same facial emotion with the same scenario. From the observer side, the same facial expression could be interpreted differently by different individuals and people often confuse one basic emotion with another (e.g., fear with surprise, Brooks et al., 2019; see also Barret et al., 2019; Elfenbein & Ambady, 2002; Durán & Fernández-Dols, 2021; for reviews). Although our LMM analysis showed that our results are less likely driven by individual differences across observers (i.e., emotion perception) or across stimuli (i.e., emotion production), whether individual differences in emotion production and perception similarly affect the perception of posed versus genuine emotional expressions remains to be established.

In conclusion, the present study investigated how stimulus-based (e.g., physical similarity) and observer-based information (e.g., perception of emotion and its intensity) contribute to perceptual similarity judgements for both static and dynamic facial expressions. Our results demonstrate that perceptual similarity between facial emotions is jointly shaped by their conceptual category, intensity, and physical similarity. These different types of information make a unique and shared contribution to perceptual similarity, with weightings varying depending on whether facial expressions are static or dynamic. These findings not only highlight novel differences between processing of static and dynamic facial emotions but also offer further evidence for the emerging view that perception of facial emotions is more nuanced and complex than that characterized by discrete emotion categories or distinctive patterns of facial movements (e.g., Brooks & Freeman 2018; Cowen & Keltner, 2021; Snoek et al. 2023).

Disclosure statement

There are no competing interests to declare.

Funding

This work is supported by a British Academy/Leverhulme Small Research Grant (SG2122\211144) awarded to MZ.

Supplementary Materials

A. Selection of stimuli from the MPI facial expression database.

Table S1. Scenarios used to elicit facial expressions of happiness, fear, and pain.

Emotion	Intensity	Scenarios
Happy	Low	*"You are lying on your couch after a delicious dinner"*
Happy	High	*"You have reached a goal, and you are happy to have accomplished it"*
Fear	Low	*"After leaving your flat you realize you forgot to switch off the cooker"*
Fear	High	*"A monster appears suddenly"*
Pain	Low	*"You watch a TV transmission of your favorite sport event. One player has a serious accident. You can see bones sticking out of the player's body"*
Pain	High	*"While doing sports you suddenly have an accident in which you twist one ankle and graze your knee."*

B. Consistency of emotion production across actors

To assess if different actors posed the same facial expressions consistently, we first extracted responses of an artificial intelligence algorithm designed to measure facial expressions (www.hume.ai), resulting in vectors along 48 emotion dimensions for each stimulus. Then, for each of the six categories of facial expressions, we measured the across-actors consistency by calculating the mean of all pairwise similarities across the nine vectors (one for each actor) using cosine similarity (i.e., 1 means identical, 0 means no similarity, and -1 means perfectly opposite). The mean across-actors consistency was 0.80 (SD: 0.07; range: 0.73-0.88 for the six facial expressions, Figure S1, left panel).

For comparison, for each actor, we also measured their within-actor consistency when they pose different facial emotions, using the mean of all pairwise cosine similarities across model responses to different categories of facial emotions (e.g., high intensity happy vs high intensity fear). The mean within-actor consistency was 0.59 (SD: 0.10; range: 0.38-0.68 for the nine actors, Figure S1, middle panel), which is significantly lower than the across-actor consistency, $t(13) = 4.51$, $p < .001$ (Figure S1, right panel). Therefore, although not all actors in the MPI facial expression database (Kaulard et al.,2012) produced identical facial movements for the same facial emotion, they did pose the same facial expressions in a highly consistent manner.

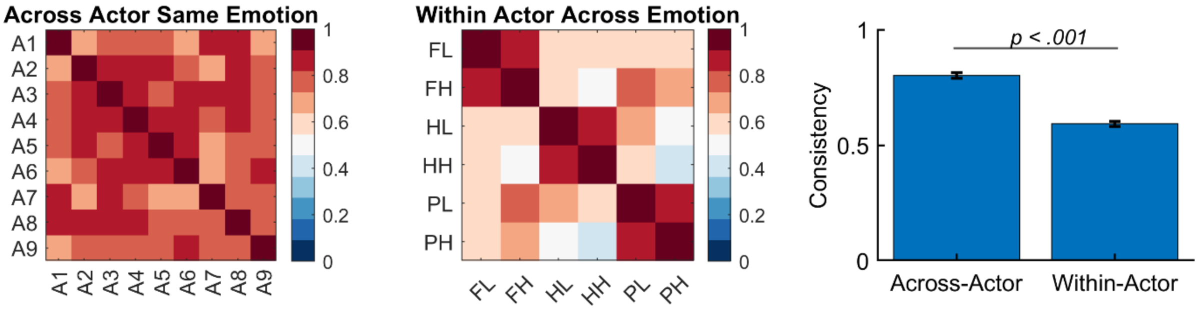

Figure S1. Consistency of emotion production across and within actors. Left, mean pairwise cosine similarity across the nine actors for the same facial expressions. Middle, mean pairwise cosine similarity across the six facial expressions within the same actor. Right, differences in mean cosine similarity between across-actor-same-expression condition and within-actor-different-emotion condition.

C. Correlation analysis across representational similarity matrices.

Table S2. Results of correlation analysis for representational similarity matrices observed in Experiment 1 (see Figure 4 in the main text for representational similarity matrices).

	Rating Static	Rating Dynamic	Categorization Static	Categorization Dynamic	Gabor Static
Rating Dynamic	.972*				
Category Static	.879*	.922*			
Category Dynamic	.869*	.925*	.992*		
Gabor Static	.935*	.883*	.715*	.715*	
Gabor Dynamic	.947*	.900*	.748*	.746*	.988*

Note. *, p <.001. *Shaded areas represent the correlations between static and dynamic conditions.*

Table S3. Results of correlation analysis for representational similarity matrices observed in Experiment 2 (see Figure 8 in the main text for representational similarity matrices).

	Rating Static	Rating Dynamic	Intensity Static	Intensity Dynamic	Gabor Static
Rating Dynamic	.966*				
Intensity Static	.906*	.848*			
Intensity Dynamic	.856*	.884*	.860*		
Gabor Static	.948*	.898*	.910*	.838*	
Gabor Dynamic	.956*	.911*	.884*	.819*	.988*

Note. *, p <.001. *Shaded area represents correlation between static and dynamic conditions*

D. Performance and comparison of regression models using all combinations of the three predictors to predict rating responses.

Table S4. Results of regression analyses using all combinations of physical (P), categorization (C), and intensity (I) similarity to predict rated similarity observed in Experiments 1 and 2.

	Predicting ratings of Experiment 1				Predicting ratings of Experiment 2			
	Static Faces		Dynamic Faces		Static Faces		Dynamic Faces	
	R^2_{Adj}	F	R^2_{Adj}	F	R^2_{Adj}	F	R^2_{Adj}	F
Model performance								
1. I	.582	222.17*	.627	267.97*	.681	337.79*	.649	294.85*
2. P	.774	545.07*	.764	515.10*	.798	622.50*	.768	527.83*
3. C	.656	304.52*	.749	474.70*	.589	229.18*	.681	340.61*
4. I+P	.780	283.21*	.806	331.37*	.826	377.78*	.818	359.13*
5. C+P	.895	680.23*	.912	823.63*	.878	571.27*	.877	565.39*
6. I+C	.802	322.08*	.858	479.96*	.822	368.61*	.824	374.19*
7. I+P+C	.897	460.79*	.926	660.57*	.897	462.20*	.899	472.18*
Model comparisons								
I+C vs I	.220	175.96*	.230	257.29*	.143	127.97*	.175	374.19*
I+C vs C	.146	116.68*	.109	121.92*	.233	207.91*	.143	129.90*
C+P vs C	.238	361.38*	.163	293.57*	.287	373.32*	.195	251.07*
C+P vs P	.121	184.02*	.148	266.52*	.082	106.72*	.108	139.68*
P+I vs P	.008	$5.58^{\dagger}$	.043	35.42*	.030	27.73*	.051	44.64*
P+I vs I	.199	143.66*	.179	147.06*	.147	133.82*	.170	148.37*
I+P+C vs P+I	.116	177.86*	.119	253.42*	.071	109.39*	.080	126.08*
I+P+C vs C+P	.002	3.17^{ns}	.014	30.01*	.020	30.36*	.023	35.72*
I+P+C vs I+C	.095	145.47*	.068	144.49*	.074	114.84*	.074	116.70*

Note. *, $p < .001$. †, $p = .019$; ns, not significant. Shaded areas show additional contribution of intensity similarity to similarity rating beyond physical and categorization similarity.*

References

Barrett, L. F. (2006). Are emotions natural kinds? *Perspectives on Psychological Science, 1(1),* 28-58. https://doi.org/10.1111/j.1745-6916.2006.00003.x

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. https://doi.org/10.1177/1529100619832930

Becker, C., Conduit, R., Chouinard, P.A., & Laycock, R. (2024). A dynamic disadvantage? Social perceptions of dynamic morphed emotions differ from videos and photos. *Journal of Nonverbal Behavior, 48*, 303–322. https://doi.org/10.1007/s10919-023-00448-3

Biele, C., & Grabowska, A. (2006). Sex differences in perception of emotion intensity in dynamic and static facial expressions. *Experimental Brain Research, 171(1),* 1-6. https://doi.org/10.1007/s00221-005-0254-0.

Bould, E., Morris, N., & Wink, B. (2008). Recognising subtle emotional expressions: The role of facial movements. *Cognition and Emotion, 22(8),* 1569-1587. https://doi.org/10.1080/02699930801921156.

Brooks, J. A., Chikazoe, J., Sadato, N., & Freeman, J. B. (2019). The neural representation of facial-emotion categories reflects conceptual structure. *Proceedings of the National Academy of Sciences of the United States of America, 116(32),* 15861-15870

Brooks, J. A., & Freeman, J. B. (2018). Conceptual knowledge predicts the representational structure of facial emotion perception. *Nature Human Behaviour, 2(8)*, 581-591. https://doi.org/10.1038/s41562-018-0376-6.

Buck, R., & VanLear, C. A. (2002). Verbal and nonverbal communication: Distinguishing symbolic, spontaneous, and pseudo-spontaneous nonverbal behavior. *Journal of Communication, 52(3),* 522-541. https://doi.org/10.1111/j.1460-2466.2002.tb02560.x

Bülthoff, I., & Zhao, M. T. (2020). Personally familiar faces: Higher precision of memory for idiosyncratic than for categorical information. *Journal of Experimental Psychology-Learning Memory and Cognition, 46(7)*, 1309-1327. https://doi.org/10.1037/xlm0000784

Bülthoff, I., & Zhao, M. (2021). Average faces: How does the averaging process change faces physically and perceptually? *Cognition, 216*, 104867. https://doi.org/10.1016/j.cognition.2021.104867

Calvo, M. G., Avero, P., Fernández-Martín, A., & Recio, G. (2016). Recognition thresholds for static and dynamic emotional faces. *Emotion, 16(8),* 1186-1200. https://doi.org/10.1037/emo0000192

Chen, C., Messinger, D. S., Chen, C., Yan, H., Duan, Y., Ince, R. A. A., Garrod, O. G. B., Schyns, P. G. & Jack, R. E. (2024). Cultural facial expressions dynamically convey emotion category and intensity information. *Current Biology, 34(1),* 213-223. https://doi: 10.1016/j.cub.2023.12.001

Cowen, A. S., & Keltner, D. (2017). Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proceedings of the National Academy of Sciences of the United States of America, 114(38),* E7900-E7909. https://doi.org/10.1073/pnas.1702247114

Cowen, A. S., & Keltner, D. (2020). What the face displays: Mapping 28 emotions conveyed by naturalistic expression. *American Psychologist, 75(3),* 349-364. https://doi.org/10.1037/amp0000488

Cowen, A. S., & Keltner, D. (2021). Semantic Space Theory: A Computational Approach to Emotion. *Trends in Cognitive Sciences, 25(2),* 124-136. https://doi.org/10.1016/j.tics.2020.11.004

Delannoy, J. R., & McDonald, J. B. (2009). Estimation of the temporal dynamics of posed and spontaneous facial expression formation using LLE. *International Machine Vision and Image Processing Conference*, 139-144. https://doi.org/10.1109/IMVIP.2009.32.

Dobs, K., Bülthoff, I., Breidt, M., Vuong, Q. C., Curio, C., & Schultz, J. (2014). Quantifying human sensitivity to spatio-temporal information in dynamic faces. *Vision Research, 100*, 78-87. https://doi.org/10.1016/j.visres.2014.04.009.

Du, S. C., Tao, Y., & Martinez, A. M. (2014). Compound facial expressions of emotion. *Proceedings of the National Academy of Sciences of the United States of America, 111(15)*, E1454-E1462. https://doi.org/10.1073/pnas.1322355111

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, 21(7)*, 1550–1569. https://doi.org/10.1037/emo0001015

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6(3-4)*, 169-200. https://doi.org/10.1080/02699939208411068

Ekman, P., & Cordaro, D. (2011). What is meant by calling emotions basic? *Emotion Review, 3(4),* 364–370. https://doi.org/10.1177/1754073911410740.

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin, 128(2)*, 203-235. https://doi.org/10.1037/0033-2909.128.2.203.

Elfenbein, H. A., Beaupré, M., Lévesque, M., & Hess, U. (2007). Toward a dialect theory: Cultural differences in the expression and recognition of posed facial expressions. *Emotion, 7(1),* 131-146. https://doi.org/10.1037/1528-3542.7.1.131.

Fiorentini, C., & Viviani, P. (2011). Is there a dynamic advantage for facial expressions? *Journal of Vision, 11(3):17*. https://doi.org/10.1167/11.3.17.

Furl, N., Begum, F., Sulik, J., Ferrarese, F. P., Jans, S., & Woolley, C. (2020). Face space representations of movement. *NeuroImage, 212*, 116676, https://doi.org/10.1016/j.neuroimage.2020.116676.

Jack, R. E., Garrod, O. G. B., Yu, H., Caldara, R., & Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences of the United States of America, 109(19),* 7241-7244. https://doi.org/10.1073/pnas.1200155109.

Jack, R. E., & Schyns, P. G. (2015). The human face as a dynamic tool for social communication. *Current Biology, 25(14),* R621-R634. https://doi.org/10.1016/j.cub.2015.05.052

Kaulard, K., Cunningham, D. W., Bülthoff, H. H., & Wallraven, C. (2012). The MPI facial expression database - a validated database of emotional and conversational facial expressions. *PLoS ONE, 7(3):* e32321. https://doi.org/10.1371/journal.pone.0032321.

Keltner, D., & Haidt, J. (1999). Social functions of emotions at four levels of analysis. *Cognition and Emotion, 13(5),* 505-521.

Krumhuber, E. G., Skora, L. I., Hill, H. C. H., and Lander, K. (2023). The role of facial movements in emotion recognition. *Nature Reviews Psychology, 2,* 283-296. https://doi.org/10.1038/s44159-023-00172-1

Krumhuber, E. G., Kappas, A., & Manstead, A. S. R. (2013). Effects of dynamic aspects of facial expressions: A review. *Emotion Review, 5(1)*, 41-46. https://doi.org/10.1177/1754073912451349.

Kuppens, P., Tuerlinckx, F., Rusell, J. A., & Barrett, L. F. (2013). The relation between valence and arousal in subjective experience. *Psychological Bulletin, 139(4),* 917-940. http://dx.doi.org/10.1037/a0030811

Lades, M., Vorbruggen, J. C., Buhmann, J., Lange, J., von der Malsburg, C., Wurtz, R. P., & Konen, W. (1993). Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers, 42(3),* 300-311. https://doi.org/10.1109/12.210173

Levenson, R. W. (2011). Basic emotion questions. *Emotion Review, 3(4),* 379-386. https://doi.org/10.1177/1754073911410743.

Lindquist, K. A., & Barrett, L. F. (2008). Constructing emotion: the experience of fear as a conceptual act. *Psychological Science, 19(9),* 898-903. https://doi.org/10.1111/j.1467-9280.2008.02174.x

Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition & Emotion, 23(2)*, 209-237. http://dx.doi.org/10.1080/ 02699930802204677

McLellan, T., Johnston, L., Dalrymple-Alford, J., & Porter, R. (2010). Sensitivity to genuine versus posed emotion specified in facial displays. *Cognition and Emotion, 24(8),* 1277–1292. https://doi.org/10.1080/02699930903306181

Mortillaro, M., Mehu, M., & Scherer, K. R. (2011). Subtly different positive emotions can be distinguished by their facial expressions. *Social Psychological and Personality Science*, 2(3), 262-271. https://doi.org/10.1177/1948550610389080.

Murray, T., O'Brien, J., Sagiv, N., & Garrido, L. (2021). The role of stimulus-based cues and conceptual information in processing facial expressions of emotion. *Cortex, 144*, 109-132. https://doi.org/10.1016/j.cortex.2021.08.007.

Muukkonen, I., & Salmela, V. R. (2022). Representational structure of fMRI/EEG responses to dynamic facial expressions. NeuroImage, 263, 119631. https://doi.org/10.1016/j.neuroimage.2022.119631.

Namba, S., Makihara, S., Kabir, R. S., Miyatani, M., & Nakao, T. (2017). Spontaneous facial expressions are different from posed facial expressions: Morphological properties and dynamic sequences. *Current Psychology, 36(3)*, 593-605. https://doi.org/10.1007/s12144-016-9448-9.

Nook, E. C., Lindquist, K. A. & Zaki, J. A. (2015). A new look at emotion perception: Concepts speed and shape facial emotion recognition. *Emotion 15(5)*, 569-578. https://doi.org/10.1037/a0039166

Panksepp, J., & Watt, D. (2011). What is basic about basic emotions? Lasting lessons from affective neuroscience. *Emotion Review, 3(4),* 387-396. https://doi.org/10.1177/1754073911410741

Pessoa, L., & Adolphs, R. (2010). Emotion processing and the amygdala: from a 'low road' to 'many roads' of evaluating biological significance. *Natural Review Neuroscience, 11,* 773-782. https://doi.org/10.1038/nrn2920

Pollick, F. E., Hill, H., Calder, A. & Paterson, H. (2003). Recognising facial expression from spatially and temporally modified movements. *Perception 32*, 813-826. https://doi.org/10.1068/p3319

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39(6),* 1161–1178. https://doi.org/10.1037/h0077714

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110(1),* 145-172. https://doi.org/10.1037/0033-295x.110.1.145

Shuman, V., Clark-Polner, E., Meuleman, B., Sander, D., & Scherer, K. R. (2015). Emotion perception from a componential perspective. *Cognition and Emotion, 31(1),* 47-56. https://doi.org/10.1080/02699931.2015.1075964.

Skerry, A. E., & Saxe, R. (2014). A common neural code for perceived and inferred emotion. *Journal of Neuroscience, 34(48),* 15997-16008. https://doi.org/10.1523/JNEUROSCI.1676-14.2014

Snoek, L., Jack, R. E., Schyns, P. G., Garrod, O. G. B., Mittenbühler, M., Chen, C., Oosterwijk, S., & Scholte, H. S. (2023). Testing, explaining, and exploring models of facial expressions of emotions. *Science Advances, 9*, eabq8421. https://doi.org/10.1126/sciadv.abq8421

Sormaz, M., Watson, D. M., Smith, W. A. P., Young, A. W., & Andrews, T. J. (2016). Modelling the perceptual similarity of facial expressions from image statistics and neural responses. *NeuroImage, 129*, 64-71. https://doi.org/10.1016/j.neuroimage.2016.01.041.

Susskind, J. M., Lee, D. H., Cusi, A., Feiman, R., Grabski, W., & Anderson, A. K. (2008). Expressing fear enhances sensory acquisition. *Nature Neuroscience, 11(7),* 843-850. https://doi.org/10.1038/nn.2138

Susskind, J. M., Littlewort, G., Bartlett, M. S., Movellan, J., & Anderson, A. K. (2007). Human and computer recognition of facial expressions of emotion. *Neuropsychologia, 45(1),* 152-162. https://doi.org/10.1016/j.neuropsychologia.2006.05.001.

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology, 66*, 519-545. https://doi.org/10.1146/annurev-psych-113011-143831

Tracy, J. L., & Randles, D. (2011). Four models of basic emotions: A review of Ekman and Cordaro, Izard, Levenson, and Panksepp and Watt. *Emotion Review, 3(4),* 397-405. https://doi.org/10.1177/1754073911410747.

Valstar, M. F., Gunes, H., & Pantic, M. (2007). How to distinguish posed from spontaneous smiles using geometric features. *Proceedings of the 9th International Conference on Multimodal interfaces*, 38-45. https://doi.org/10.1145/1322192.1322202

Van Kleef, G. A. (2009). How emotions regulate social life: The emotions as social information (EASI) model. *Current Directions in Psychological Science, 18(3),* 184-188. https://doi.org/10.1111/j.1467-8721.2009.01633.x.

Williams, A. C. D. (2002). Facial expression of pain: An evolutionary account. *Behavioral and Brain Sciences, 25(4),* 439-488. https://doi.org/Doi 10.1017/S0140525x02000080

Xu, X., & Biederman, I. (2010). Loci of the release from fMRI adaptation for changes in facial expression, identity, and viewpoint. *Journal of Vision, 10(14):* 36. https://doi.org/10.1167/10.14.36.

Yue, X., Biederman, I., Mangini, M. C., Malsburg, C. von der, & Amir, O. (2012). Predicting the psychophysical similarity of faces and non-face complex shapes by image-based measures. *Vision Research, 55*, 41-46. https://doi.org/10.1016/j.visres.2011.12.012.