Acta Psychologica 187 (2018) 19–29

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

Discrimination between smiling faces: Human observers vs. automated face analysis

Mario Del Líbano[a], Manuel G. Calvo[b,*], Andrés Fernández-Martín[c], Guillermo Recio[d]

[a] Universidad de Burgos, Burgos, Spain
[b] Universidad de La Laguna, Tenerife, Spain
[c] Universidad Internacional de La Rioja, Logroño, Spain
[d] Universität Hamburg, Hamburg, Germany

ARTICLE INFO

Keywords:
Facial expression
Smile
Emotion
Action units
FACET

ABSTRACT

This study investigated (a) how prototypical happy faces (with happy eyes and a smile) can be discriminated from blended expressions with a smile but non-happy eyes, depending on type and intensity of the eye expression; and (b) how smile discrimination differs for human perceivers versus automated face analysis, depending on affective valence and morphological facial features. Human observers categorized faces as happy or non-happy, or rated their valence. Automated analysis (FACET software) computed seven expressions (including joy/happiness) and 20 facial action units (AUs). Physical properties (low-level image statistics and visual saliency) of the face stimuli were controlled. Results revealed, first, that some blended expressions (especially, with angry eyes) had lower discrimination thresholds (i.e., they were identified as "non-happy" at lower non-happy eye intensities) than others (especially, with neutral eyes). Second, discrimination sensitivity was better for human perceivers than for automated FACET analysis. As an additional finding, affective valence predicted human discrimination performance, whereas morphological AUs predicted FACET discrimination. FACET can be a valid tool for categorizing prototypical expressions, but is currently more limited than human observers for discrimination of blended expressions. Configural processing facilitates detection of in/congruence(s) across regions, and thus detection of non-genuine smiling faces (due to non-happy eyes).

1. Introduction

A smile (basically, lip corners turned up and pulled backwards, frequently accompanied by exposed upper teeth) is often assumed to be a diagnostic facial feature of happiness and to reflect positive feelings and motives. Because people typically smile when they are happy, observers generally infer that the smiler feels happy (and/or is probably friendly). However, being happy does not necessarily lead someone to smile, and a smile can be exhibited for reasons unrelated to happiness. Actually, the smile is multifaceted and multifunctional in social interaction (Ambadar, Cohn, & Reed, 2009; Crivelli, Carrera, & Fernández-Dols, 2015; Ekman, 2001; Niedenthal, Mermillod, Maringer, & Hess, 2010). In addition to reflecting positive feelings (enjoyment, warmth, etc.), smiles can conceal or leak non-positive feelings, motives, or intentions (mockery, contempt, arrogance, malicious joy or schadenfreude, embarrassment, nervousness, etc.), or portray mere social politeness devoid of affect. Further, a person can involuntarily experience mixed emotions simultaneously, even involving opposite feelings of pleasantness and unpleasantness (see Russell, 2017), which can produce a variety of blended facial expressions with a smile. Thus, it is important to identify and differentiate the significance of such a variety of smiles.

For an observer, to distinguish a smile conveying positive feelings, motives, and intentions from a smile lacking them (or concealing non-positive ones), contextual factors and prior knowledge of the expresser can play an important role (Fernández-Dols & Crivelli, 2013; Hassin, Aviezer, & Bentin, 2013). In addition, a morphological facial feature called the Duchenne or D marker in the eye region can make a significant contribution (see Gunnery & Ruben, 2016). This marker engages contraction of the *orbicularis oculi* muscle, which lifts the cheek, narrows the eye opening, and produces wrinkles around the eyes (Frank, Ekman, & Friesen, 1993). Although such a marker can be spontaneous or deliberate (Krumhuber & Manstead, 2009), its absence or replacement with negatively valenced expressive changes (e.g., frown, etc.) would indicate that the smile does not reflect authentic happiness. A recent meta-analysis (Gunnery & Ruben, 2016) has shown

* Corresponding author at: Department of Cognitive Psychology, University of La Laguna, 38205 Tenerife, Spain.
E-mail addresses: mdlibano@ubu.es (M. Del Líbano), mgcalvo@ull.edu.es (M.G. Calvo), andres.fernandez@unir.net (A. Fernández-Martín), guillermo.recio@uni-hamburg.de (G. Recio).

https://doi.org/10.1016/j.actpsy.2018.04.019
Received 27 November 2017; Received in revised form 9 April 2018; Accepted 30 April 2018
0001-6918/

that Duchenne smiles and people producing them are rated more positively (i.e., authentic, real, attractive, trustworthy) than non-Duchenne smiles. Thus, observers can to some extent discriminate smiles assumed to convey positive affect from other smiles by relying on the eye region expression (see Ambadar et al., 2009; Gunnery & Hall, 2014; Krumhuber, Likowski, & Weyers, 2014; McLellan, Johnston, Dalrymple-Alford, & Porter, 2010; McLellan, Wilcke, Johnston, Watts, & Miles, 2012; Miles & Johnston, 2007; Quadflieg, Vermeulen, & Rossion, 2013; Slessor et al., 2014).

However, discrimination is limited and sometimes smiling faces with non-D eyes are seen as if they showed genuine happiness (see Krumhuber et al., 2014; Okubo, Kobayashi, & Ishikawa, 2012; Quadflieg et al., 2013). A central question is the influence of upper-face action and intensity in modifying the perceived meaning of smiles. In this context, to examine the role of the eye expression in discriminating among different types of smiling faces, a series of studies have been conducted in which the *type* of non-happy eye expression was varied. In addition to explicit expression recognition, measures of affective priming (Calvo, Fernández-Martín, & Nummenmaa, 2012), eye movements (Calvo, Gutiérrez-García, Avero, & Lundqvist, 2013), event-related potentials (ERPs) of brain activity (Calvo, Marrero, & Beltrán, 2013), and perceptual thresholds (Gutiérrez-García & Calvo, 2015) were collected, using blended expressions (i.e., a smiling mouth but non-happy eyes—neutral, angry, fearful, etc.) as stimuli. For comparison, prototypical happy faces (smiling mouth and Duchenne, happy eyes) and prototypical non-happy faces (e.g., angry mouth and angry eyes, etc.) were also presented. Across the various paradigms, difficulties in identifying (as "non-happy") blended expressions with a smile increased in the presence of angry vs. disgusted vs. sad vs. fearful vs. surprised or neutral eyes. That is, discrimination was better for smiling faces with angry eyes (i.e., the least likely to be confused as happy) than for those with disgusted eyes, which were discriminated better than those with sad or fearful eyes, and discrimination was poorest for smiling faces with surprised or neutral eyes.

The current study extended prior research with two major aims. First, we investigated smile discrimination *thresholds* and gradients, depending on type and intensity of the eye expression. We determined threshold as the *minimum* expressive intensity of happiness in the eye region that is required to recognize a smile as conveying positive feelings, as well as the minimum intensity of non-happy eye expressions (i.e., angry, etc.) that allows observers to identify a smiling face as *not* truly happy. To this end, using Karolinska Directed Emotional Faces (KDEF; Lundqvist, Flykt, & Öhman, 1998) as stimuli, (a) we varied the degree of intensity (11 levels, from 0 to 100%, in 10% steps) of different eye expressions (happy, angry, fearful, disgusted, sad, surprised, or neutral) by means of a graphics morphing software (FantaMorph; Abrosoft; see 2.2. Stimuli); and (b) we combined a smiling mouth with different eye regions by means of the composite face technique (e.g., Quadflieg et al., 2013; Tanaka, Kaiser, Butler, & Le Grand, 2012). Thus, we generated photographs of (a) *prototypical* happy expressions with a smiling mouth and a happy eye expression, and (b) *blended* expressions with the same smiling mouth, but an eye expression that varied in intensity (from happy to non-happy).

The second aim addressed the comparison of *human* 'subjective' perception vs. *automated* 'objective' assessment of smile discrimination. An important issue here is the relative role of affective valence (as measured by subjective ratings) vs. physical features (as measured by objective automated analysis) of facial expressions. To this end, from human observers, we obtained (a) the probability that they categorized faces as happy or not happy; and (b) the degree of affective valence rating of each face, i.e., how positive or negative the expression configuration looked like. In addition, by means of automated analysis with Emotient FACET software (iMotions; see 2.5. Automated analysis of facial expressions), we computed (a) the probability of each of the six basic emotions (joy, anger, etc.) and neutrality for each face stimulus; and (b) each of 20 morphological action units (AUs) at local regions. Prior assessment of AUs by the Facial Action Coding System (FACS; Ekman & Friesen, 1978; Ekman, Friesen, & Hager, 2002) has shown that specific muscle movements characterize different emotional expressions. Recent developments (e.g., FACET) have standardized the assessment, allowing for a quantification of emotions and AUs as a function of spatial parameter maps of facial features (see Bartlett & Whitehill, 2011; Cohn & De la Torre, 2015).

In a related approach, Calvo, Gutiérrez-García, and Del Líbano (2018) recently found that, for human observers, (a) the probability of perceiving happiness in prototypical happy faces and also in blended expressions with a smile increased mainly as a function of affective valence of the facial configuration; and (b) the probability of (wrongly) perceiving blended expressions as happy increased with delayed saliency and reduced distinctiveness of the non-happy eye region, and with enhanced AU6 (cheek raiser) and reduced AU4 (brow lowerer). The current design makes three significant contributions. First, we have directly compared human processing and automated modelling of prototypical happy faces and blended expressions with a smile. Second, by means of automated analysis, we have assessed 20 AUs, and also six other expressions apart from happiness. Third, the eye expression intensity has now been systematically varied to examine discrimination thresholds, whereas previously only single (apex) happy or non-happy eye expressions were presented.

For the current study, we conducted two experiments, with two samples of 100 participants each and two different tasks, either happiness categorization or affective valence ratings. In addition, we performed computational modelling of facial expressions and AUs, as well as assessment of physical properties of the face stimuli (low-level image statistics and visual saliency). Nevertheless, for economy of exposition, and to provide an integrated view of the different measures, all of them will be presented together (in the Materials and methods, Results, and Discussion sections), as parts of the same study.

2. Materials and methods

2.1. Participants

Two-hundred university undergraduates (124 female; 76 male; aged 18 to 30 years; M = 21.1 years) from different courses (Psychology, Medicine, Law, Economics, and Education) participated voluntarily or for course credit, after providing informed consent. One-hundred of them (62 female) were randomly assigned to a facial happiness judgment task, and another 100 (62 female), to a valence rating task (see 2.4. Procedure). The study was approved by the University of La Laguna Ethics Committee, and was conducted in accordance with the WMA Declaration of Helsinki 2008.

2.2. Stimuli

For the different tasks (happiness judgment, valence rating, automated assessment, and computation of physical image properties), we used color photographs from the KDEF set (Lundqvist et al., 1998). The face stimuli portrayed 24 individuals (12 females: KDEF model numbers 01, 02, 07, 11, 14, 19, 20, 22, 26, 29, 31, 33; 12 males: 03, 05, 06, 10, 11, 12, 22, 23, 24, 25, 31, 35, each posing seven facial expressions (neutral, happiness, anger, disgust, sadness, fear, and surprise). All 24 models were presented once in their original form as (a) prototypical *happy* expressions showing Duchenne eyes and a smile, and as (b) prototypical *non-happy* expressions (neutral, anger, etc.).

In addition, (c) based on the KDEF original stimuli, we constructed six *blended* expressions with a smile but non-happy eyes, thus producing 144 new face stimuli, by means of the composite face technique (e.g., Tanaka et al., 2012). The upper half of each non-happy face and the lower half of the happy face were combined, by cutting each face along a horizontal line through the bridge of the nose and smoothing the junction. The following blends were created for each of the 24 models:

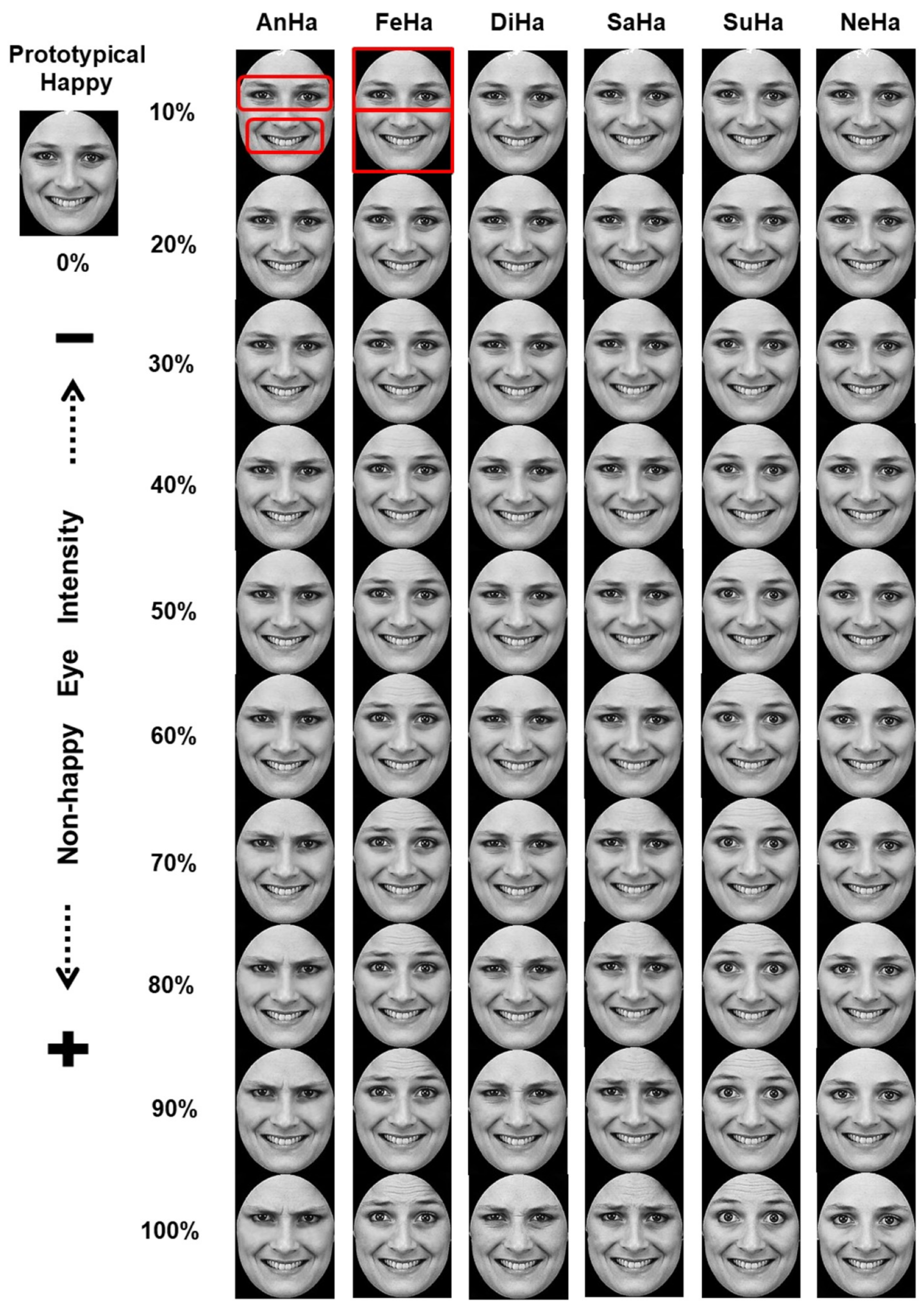

Fig. 1. Sample stimuli. Examples of prototypical happy faces (happy eyes and a smile) and blended expressions (non-happy eyes and a smile) across 10% intensity levels for each type of non-happy eyes. AnHa: angry eyes with a smile; FeHa: fearful eyes with a smile; DiHa: disgusted eyes with a smile; SaHa: sad eyes with a smile; SuHa: surprised eyes with a smile; NeHa: neutral eyes with a smile. Top left corner (AnHa and FeHa, 10% intensity): Face areas (eyes, mouth, upper half, lower half, whole face) on which low-level image statistics (i.e., luminance, etc.) and visual saliency were computed for all the stimuli.

Neutral eyes + Happy smile (NeHa), Angry eyes + Happy smile (AnHa), Disgusted eyes + Happy smile (DiHa), Sad eyes + Happy smile (SaHa), Fearful eyes + Happy smile (FeHa), and Surprised eyes + Happy smile (SuHa) (see Fig. 1).

Finally, (d) the expressive *intensity* of the *non-happy eyes* (e.g., angry) in the *blended* expressions was morphed by means of FantaMorph© software (v.5.4.2; Abrosoft, Beijing, China). For each expression of each poser, we created a sequence of 30 frames based on two images: (1) a prototypical happy face (with happy eyes and a smile), as the first frame, and (2) a blended expression with the eyes of a prototypical non-happy face (angry, etc.), but the smile of the prototypical happy face, as the final frame. We selected the first frame (i.e., the prototypical happy face, with 100% intensity of happy eyes; hence 0% intensity of non-happy eyes) and 10 additional frames that represented, respectively, the 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100% intensities of each non-happy eye expressions (neutral, anger, disgust, sadness, fear, and surprise; see Fig. 1).

In sum, we used 1440 photographs (24 posers × 6 non-happy eye expressions × 10 intensities) of *blended* expressions (a smiling mouth with non-happy eyes), in addition to 24 photographs (24 posers) of prototypical *happy* (both eye and mouth, 100% intensities) expressions, and 144 photographs of prototypical *non-happy* (both eye and mouth, 100% intensities) expressions (24 posers × 6 non-happy—angry, sad, fearful, disgusted, surprised, and neutral—faces).

2.3. Prototypical happy vs. non-happy smiling faces

To ensure that our prototypical happy face stimuli conveyed “truly” happy expressions, we used three criteria. First, we chose KDEF models that included AU6 (i.e., cheek raiser, with the D-marker around the eye region) in addition to AU12 (i.e., lip corner puller), as measured by FACET (see 2.6. Automated analysis of facial expressions). Prototypical happy faces showed AU12 to a greater extent than all the prototypical non-happy faces, $F(6, 161) = 143.65, p < 0.0001, \eta_p^2 = 0.84$, and also higher AU6 scores, $F(6, 161) = 93.98, p < 0.0001, \eta_p^2 = 0.78$ (see Table 2 of Supplemental Analyses and Tables). Second, in a different study (Calvo & Fernández-Martín, 2013), the face upper-half alone was presented for 150 ms, and participants responded whether the eye expression was happy or not. The eye region of the prototypical happy faces was correctly identified as happy by 81% of participants. Third, the prototypical happy faces have been found to produce affective priming on the processing of pleasant scenes, whereas blended expressions with the same smiling mouth but non-happy eyes did not produce affective priming (Calvo et al., 2012). Altogether, this allows us to argue that our prototypical happy faces displayed "truly" happy smiles.

2.4. Procedure

In the *happiness categorization* task and the *affective valence* task (100 different participants each), each observer was presented with 192 photographs as experimental stimuli: 24 prototypical happy faces (24 KDEF posers) and 144 blended expressions with a smile but non-happy eyes, in addition to 24 prototypical non-happy faces (four KDEF posers of each of the six non-happy expressions: 4 angry, 4 neutral, etc.) that served as fillers. To avoid repetitions of the same poser and expression at different intensities to the same participant, the stimuli were combined into 10 different counterbalancings (with each one being assigned to 10 different participants). Thus, for blended expressions, each participant was presented with each poser six times, albeit only once in each of the six expressions, and each time at a different intensity level. Each prototypical happy face was rated by all the participants. All six categories of blended expressions were seen by all the participants, with each poser being rated by 10 participants in each intensity condition of each blended expression. The stimuli were presented in four blocks (counterbalanced order) of 48 trials (in randomized sequence) each, following 26 practice trials.

The stimuli were displayed on a computer screen using E-Prime 2.0 software (Schneider, Eschmann, & Zuccolotto, 2002). Participants were informed that faces would be presented with different expressions (otherwise unspecified). In the *happiness categorization* task, participants indicated "whether a face looked happy or not happy", by pressing a "Yes" (L, in a computer keyboard) or "No" (D) key with their left or right forefingers, as soon as possible. In the *affective valence* task, participants rated "how emotionally negative or positive" the facial expression was, using the standard Self-Assessment Manikin on a 1–9 scale (Lang, Bradley, & Cuthbert, 2008), by pressing one out of 9 keys on the upper row of the keyboard.

The sequence of events on each trial was similar for both tasks. After an initial 500-ms central fixation cross on a screen, a face stimulus appeared until the participant responded. The face subtended a visual angle of 10.6° (height) × 8° (width) at a 70-cm viewing distance (this approximates the size of a real face, i.e., 18.5 × 13.8 cm, viewed from a 1-m distance). The selected response and reaction times were collected for the happiness task, while only the selected response was used for the valence task (as reaction times might be influenced by the different location of the keys for the 1–9 scale).

2.5. Automated analysis of facial expressions

The face stimuli were subjected to automated expression analysis (see Bartlett & Whitehill, 2011; Cohn & De la Torre, 2015; Girard, Cohn, Jeni, Sayette, & De la Torre, 2015; Gordon, Tanaka, Pierce, & Bartlett, 2011; Olderbak, Hildebrandt, Pinkpank, Sommer, & Wilhelm, 2014) by means of Emotient FACET SDK v6.1 (iMotions, A/S, Copenhagen, Denmark), which provides two measures. First, so-called *expression* evidence scores quantify the probability of each expression category to be present in a given face image (as labeled in FACET: joy, anger, surprise, fear, disgust, sadness, and contempt, in addition to neutral). Second, *action unit* (AU) evidence scores quantify several AUs (AUs 1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 18, 20, 23, 24, 25, 26, 28, and 43), according to the Facial Action Coding System (FACS; and Facial Action Coding System Affect Interpretation Database, FACSAID; Ekman et al., 2002; see also Cohn, Ambadar, & Ekman, 2007). AUs are anatomically related to the movement of specific face muscles (e.g., AU12 involves contraction of the zygomaticus major muscle, which draws the angle of the mouth superiorly and posteriorly to allow for smiling). FACET evidence scores are expressed in odds ratios in decimal logarithmic scale, with positive values indicating that an expression or AU are present; negative values, that they are not present; a zero score indicates chance level (0.50/0.50).

FACET is assumed to classify face images as expressions by relying on the detection of morphological cues. This assumption was supported in the current study, as expressions were associated with specific AUs (see 3.3.2. Discrimination of AUs). Early work showed that automated facial expression analysis can discriminate dynamic facial configurations differing in psychological meaning (e.g., spontaneous vs. deliberate smiles: Cohn & Schmidt, 2004; or spontaneous vs. posed brow actions: Valstar, Pantic, Ambadar, & Cohn, 2006). This led us to use this type of computational modelling to discriminate smiles in prototypical vs. blended expressions.

2.6. Low-level image properties and visual saliency

As control measures, we assessed low-level image statistics and visual saliency of the face stimuli at each intensity level. This served to examine whether discrimination thresholds (as a function of type and intensity of the eye expression in smiling faces) would be confounded with purely physical properties of the face images. First, with Matlab 7.0 (The Mathworks, Natick, MA), we computed luminance, root-mean-square (RMS) contrast, skewness, kurtosis, energy, and signal-to-noise (SNR) ratio. These measures were obtained from the whole face, the upper and the lower face half, and also for the *eye region*—which was particularly important, as this region (and hence also the upper face half) varied for the different blended expressions and expressive intensities—and the mouth region, separately (see Fig. 1). Second, with the iLab Neuromorphic Vision C+ Toolkit (iNVT; http://ilab.usc.edu/toolkit; Itti & Koch, 2000; Walther & Koch, 2006), we computed the visual saliency of the eye and the mouth regions (see Fig. 1). This algorithm calculates the visual conspicuity of an image area (pixel-by-pixel) as a function of a combination of contrast, color, and spatial orientation. Unlike the low-level image statistics, which are measured for each region regardless of the rest of the face, visual salience represents the *relative* weight of each region within the whole face. Visual saliency has proved to influence shifts of covert and overt attention (see Borji & Itti, 2013; Calvo & Nummenmaa, 2008; Underwood & Foulsham, 2006).

3. Results

Given that automated facial expression assessment had to be computed for each face stimulus, and that we wanted to compare human observers and automated modelling, the face stimuli were the units of statistical analysis. Type of eye expression (happy, angry, disgusted, sad, fearful, surprised, and neutral) and expressive intensity (from 0% to 100%, in 10% steps) of the non-happy eyes in the blended expressions were repeated-measures factors. For all post hoc multiple comparisons, Bonferroni corrections ($p < 0.05$) were used. The raw data are included in a Supplemental Dataset.

3.1. Low-level image properties and visual saliency of the face stimuli

A 6 (Non-happy eye Expression: angry, disgusted, fearful, sad, surprised, and neutral) × 11 (eye Intensity: 0% non-happy [i.e., 100% happy], 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100% non-happy) ANOVA was conducted on each low-level image property and the visual saliency measures. This served to examine whether changes in such properties as a function of intensity differed across the various blended expressions with non-happy eyes relative to prototypical happy faces. No significant interaction emerged for any measure or face area (all *F*s < 1; *ps* ≥ 0.90, *ns*). This implies that changes in such physical properties as a function of intensity were comparable for the different non-happy eye expressions. Thus, potential differences in discrimination thresholds across expressions (see Section 3.2.) would not be due to confounding physical stimulus factors.

3.2. Human observers' performance

To establish discrimination thresholds in the *happiness categorization* task, we used two complementary criteria. A stringent criterion determined the lowest expressive intensity level at which each blended expression was judged as "*non-happy*". A lenient criterion determined the lowest intensity at which the probability of judging a blended expression as "happy" was significantly lower (i.e., *less happy*, which also implies discrimination) than for the prototypical happy face. In addition, the asymptote, i.e., the intensity at which response *latencies* no longer increased, and started to decline, were assumed to reflect facilitated discrimination of blended expressions as "non-happy". Relatedly, analyses of *affective valence* ratings determined the lowest intensity at which the valence of each blended expression was lower than for the prototypical happy faces.

3.2.1. Smile discrimination thresholds and asymptotic response latencies

A 6 (Non-happy eye Expression) × 11 (eye Intensity) ANOVA was conducted on the probability of judging faces as happy. Main effects of eye expression, $F(5, 1518) = 586.68$, $p < 0.0001$, $\eta_p^2 = 0.66$, and intensity, $F(10, 1518) = 1843$, $p < 0.0001$, $\eta_p^2 = 0.92$, were qualified by an interaction, $F(50, 1518) = 19.93$, $p < 0.0001$, $\eta_p^2 = 0.40$. To decompose the interaction, one-way (Intensity) ANOVAs were performed for each expression. The critical comparisons involved the prototypical happy faces vs. each intensity level of the blended expression. As indicated in Fig. 2 (upper half), the discrimination threshold was located at 20% intensity for AnHa (angry eyes) faces; 30%, for FeHa (fearful eyes), DiHa (disgusted eyes), and SaHa (sad eyes) faces; 40%, for SuHa (surprised eyes) faces; and 50%, for NeHa (neutral eyes) faces; all $Fs(10, 253) \geq 146.80$, $p < 0.0001$, $\eta_p^2 \geq 0.85$. Thus, AnHa faces needed less intensity, and NeHa faces needed more intensity, to be discriminated from prototypical happy expressions, relative to the other blends.

The previous analysis revealed the threshold at which blended expressions were judged as "*different*" from (i.e., less happy than) the prototypical happy expressions, but the rating probabilities (as "happy") were still high (~0.80 or above). To examine when each blended expression was judged as "*non-happy*" (i.e., more non-happy than happy), one-sample *t*-tests were conducted against the 0.5 probability. As indicated in Fig. 2 (lower half), AnHa faces were considered as non-happy at 60% intensity (of the non-happy eyes) and above, $t(23) = 5.80$, $p < 0.0001$, $d = 1.88$; FeHa, $t(23) = 2.30$, $p = 0.03$, $d = 0.92$, DiHa, $t(23) = 3.68$, $p = 0.001$, $d = 1.51$, and SaHa, $t(23) = 3.11$, $p = 0.005$, $d = 1.24$, faces, at 70% intensity; and SuHa, $t(23) = 8.14$, $p < 0.0001$, $d = 1.88$, and NeHa, $t(23) = 3.41$, $p = 0.002$, $d = 1.39$, faces, only at 100% intensity. Thus, comparable patterns across expressions appeared for the "different" and the "non-happy" discrimination thresholds.

A 6 (Expression) × 11 (Intensity) ANOVA on *response latencies* yielded effects of eye expression, $F(5, 1518) = 19.19$, $p < 0.0001$, $\eta_p^2 = 0.06$, intensity, $F(10, 1518) = 230.05$, $p < 0.0001$, $\eta_p^2 = 0.60$, and an interaction, $F(50, 1518) = 10.47$, $p < 0.0001$, $\eta_p^2 = 0.26$. Follow-up one-way (Intensity) ANOVAs were conducted for each expression, all $Fs(10, 253) \geq 16.90$, $p < 0.0001$, $\eta_p^2 \geq 0.40$. The

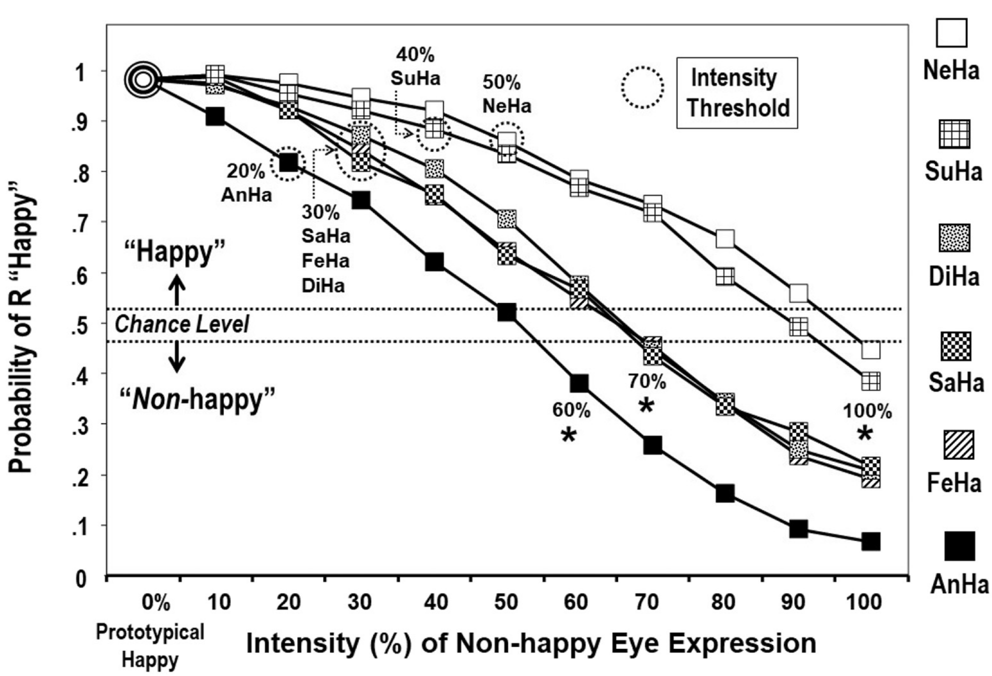

Fig. 2. Categorization of faces as happy. Mean probability of human observers judging a face as "happy" for prototypical happy expressions (happy eyes and a smile) and blended expressions (non-happy eyes and a smile) across 10% intensity levels for each type of non-happy eyes. Dotted circles (upper half of figure) indicate discrimination thresholds (i.e., lowest intensity at which blended expressions were perceived as *less happy* than prototypical happy faces). Asterisks (lower half of figure) indicate the thresholds at which blended expressions were perceived as *non-happy*. For acronyms (AnHa, etc.), see Fig. 1 caption.

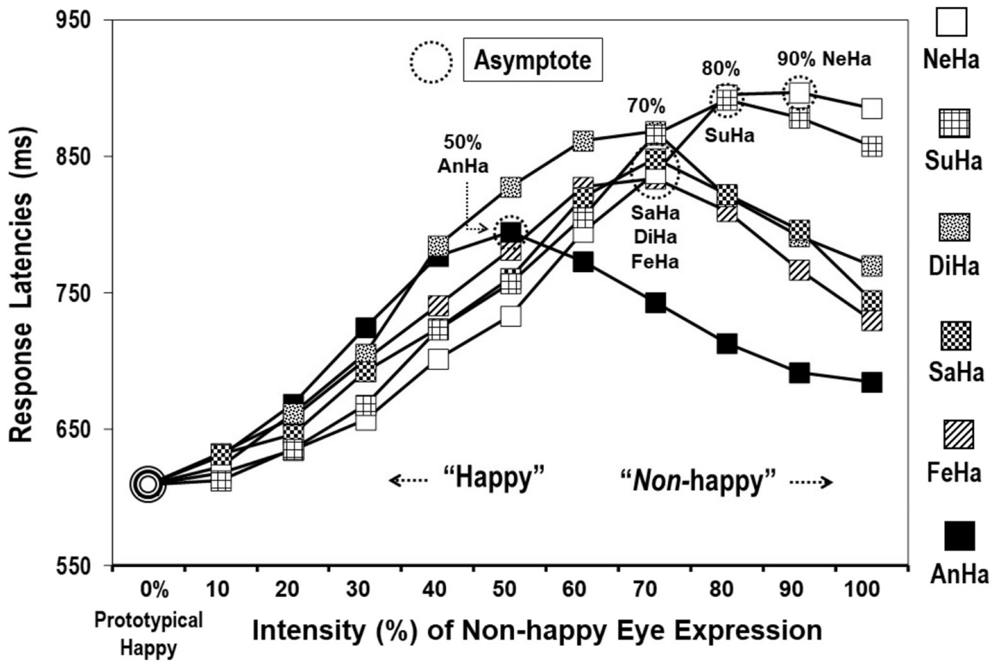

Fig. 3. Response latencies. Mean reaction times of human observers judging face stimuli as “happy” or “non-happy” for prototypical happy expressions (happy eyes and a smile) and blended expressions (non-happy eyes and a smile) across 10% intensity levels for each type of non-happy eyes. Dotted circles indicate asymptotic levels.

interaction revealed that the asymptote was reached at lower intensities for some expressions than for others (see Fig. 3): AnHa faces, 50% intensity; FeHa, DiHa, and SaHa faces, 70%; SuHa faces, 80%; and NeHa faces, 90%. This indicates that discrimination was most demanding of processing resources for NeHa faces, and most efficient for AnHa faces, relative to the other blended expressions.

3.2.2. *Affective valence thresholds*

The 6 (Expression) × 11 (Intensity) ANOVA on affective valence ratings showed effects of expression, $F(5, 1518) = 503.60, p < 0.0001$, $\eta_p^2 = 0.62$, intensity, $F(10, 1518) = 1734$, $p < 0.0001$, $\eta_p^2 = 0.92$, and an interaction, $F(50, 1518) = 38.90, p < 0.0001$, $\eta_p^2 = 0.56$. One-way (Intensity) ANOVAs revealed the minimum intensity at which affective valence of each blended expression was significantly lower than for prototypical happy expressions (see Fig. 4): The threshold was located at 20% intensity for AnHa (angry eyes) faces; 30% for DiHa (disgusted eyes) and SaHa (sad eyes) faces; and 40% for FeHa (fearful eyes), SuHa (surprised eyes), and NeHa (neutral eyes) faces; all $Fs(10, 253) \geq 96.46, p < 0.0001$, $\eta_p^2 \geq 0.79$. Thus, AnHa faces conveyed less positive affect from lower intensities, relative to the other blended expressions.

In addition, a one-way (6: Blended expression) ANOVA was conducted on the *difference* in affective valence scores between each blended expression (at 100% intensity of non-happy eyes) and the prototypical happy expression (i.e., blended minus happy). This served to determine how much affective valence decreased as a function of type of non-happy eyes. A main effect, $F(5, 138) = 154.15$, $p < 0.0001$, $\eta_p^2 = 0.85$, and post hoc contrasts (all $ps < 0.05$) revealed the greatest reduction in ratings of positive valence occurred for AnHa faces (M difference = −5.09), followed by DiHa (−3.84) faces,

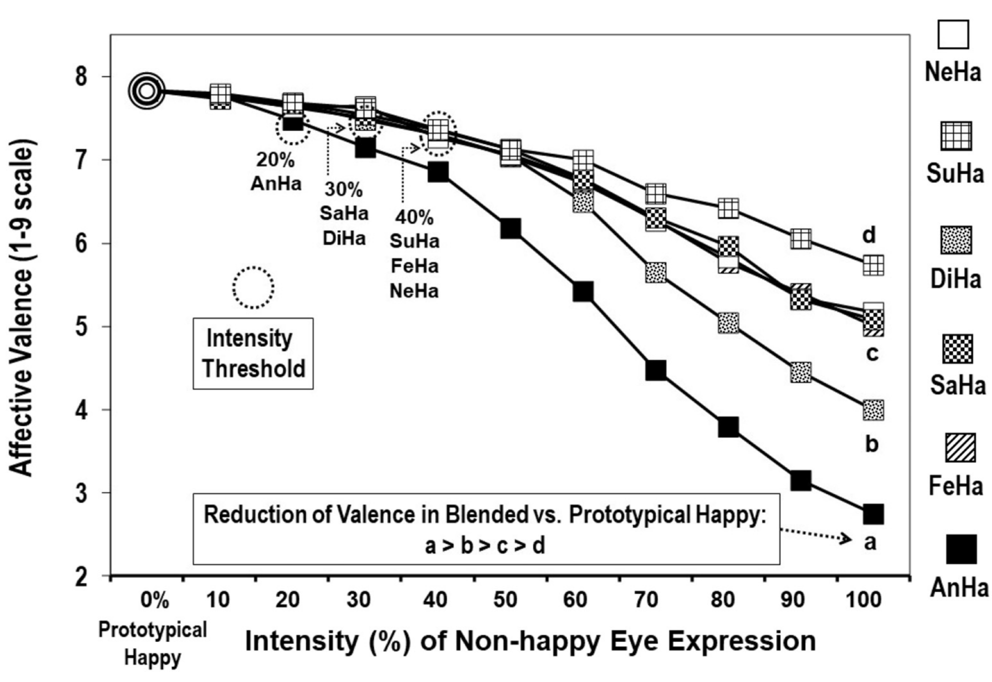

Fig. 4. Affective valence ratings. Mean affective valence ratings of human observers for prototypical happy expressions (happy eyes and a smile) and blended expressions (non-happy eyes and a smile) across 10% intensity levels for each type of non-happy eyes. Dotted circles indicate discrimination thresholds. Letters (a to d) indicate significant differences among blended expressions. > Greater reduction from prototypical to blended expression (e.g., for a [AnHa] vs. b [DiHa], etc.).

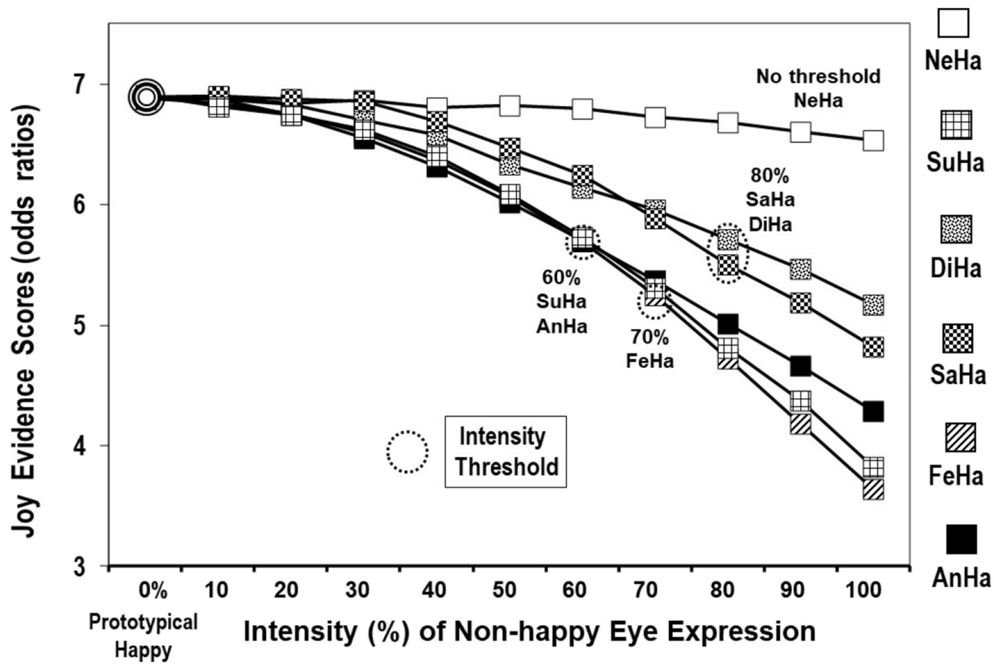

Fig. 5. Automated assessment of happiness. Mean joy evidence scores from automated assessment for prototypical happy expressions (happy eyes and a smile) and blended expressions (non-happy eyes and a smile) across 10% intensity levels for each type of non-happy eyes. Dotted circles indicate discrimination thresholds.

followed by FeHa (−2.83), SaHa (−2.76), and NeHa (−2.66) faces (which did not differ significantly from one another), and the smallest reduction was for SuHa (−2.09) faces (see Fig. 4).

3.3. Automated facial expression assessment

The specific aim of this study regarding automated facial expression assessment was concerned with discrimination between prototypical happy expressions and blended expressions with a smile. Nevertheless, a requisite was that automated assessment was *sensitive* to *prototypical* expressions and the *AUs* associated with them according to the Facial Action Coding System. To this end, we conducted *specificity checks*. The results are presented as Supplemental Analyses and Tables: Essentially, FACET software showed satisfactory classification of prototypical expressions and AUs.

Next, we report the analyses of results regarding automated expression coding and AUs for *blended* vs. *happy* expressions. Analyses were conducted to determine, first, the minimum expressive intensity (i.e., the threshold) at which the classification of a blended expression as "*joy*" was significantly *lower* (thus implying discrimination) than for the prototypical happy expression; and, second, the lowest intensity at which each AU differed for blended relative to prototypical happy expressions.

3.3.1. Smile discrimination: joy evidence

A 6 (Expression) × 11 (Intensity) ANOVA was conducted on *joy* evidence scores, as computed by FACET. Effects of eye expression, $F(5, 1518) = 33.74$, $p < 0.0001$, $\eta_p^2 = 0.10$, intensity, $F(10, 1518) = 69.27$, $p < 0.0001$, $\eta_p^2 = 0.31$, and an interaction, $F(50, 1518) = 2.82$, $p < 0.0001$, $\eta_p^2 = 0.09$, emerged. In follow-up one-way (Intensity) ANOVAs, the critical comparisons involved the prototypical happy faces vs. each blended expression. As indicated in Fig. 5, the threshold was at 60% intensity for AnHa (angry eyes) and SuHa (surprised eyes) faces; 70% for FeHa (fearful eyes) faces; 80% for DiHa (disgusted eyes) and SaHa (sad eyes) faces; all $Fs(10, 253) \geq 8.76$, $p < 0.0001$, $\eta_p^2 \geq 0.26$. AnHa and SuHa faces were thus discriminated at lower intensities than the other blended expressions. In contrast, NeHa (neutral eyes) faces were not discriminated from prototypical happy faces at any intensity level.

3.3.2. Discrimination of AUs

One-way (11 non-happy eye Intensity) ANOVAs were conducted on each AU scores. Out of the 20 AUs measured by FACET, 10 showed statistically significant differences between the prototypical happy faces and at least one or more blended expressions (see Table 1). Such AUs—which were all FACS-relevant for the different emotions—were *retained* for the subsequent analyses. Mean scores and intensity thresholds for each AU are shown in Table 1. For example, AnHa faces differed from prototypical happy faces on AU1 (at 70% intensity), AU2 (50% intensity), etc., and AU23 (60%), all $Fs(10, 253) \geq 2.98$, $p < 0.001$, $\eta_p^2 \geq 0.10$. The other blended expressions, except NeHa faces, also differed significantly on several AUs at various intensities, all $Fs(10, 253) \geq 3.90$, $p < 0.0001$, $\eta_p^2 \geq 0.13$ (see details in Table 1).

3.4. Relationships between human and software performance

To examine the relationship between human and automated smile discrimination, we conducted two analyses. First, we performed Pearson correlations between the human observers' measures (i.e., probability of judging a face as happy, and affective valence ratings) and the automated assessment measures (i.e., probability of classifying a face as conveying joy, and the 10 selected AUs). As shown in Table 2, the probability that human observers judged faces as happy was correlated with the automated joy evidence scores, $r(1584) = 0.54$, $p < 0.0001$, 95% CI [0.50, 0.57]. Intra-class correlation (ICC 2, two-way random) analyses revealed reliable classification consistency between evidence scores and human ratings of faces as happy or non-happy, for each blended expression ($N = 264$; AnHa: $r = 0.65$, $p < 0.0001$; FeHa: $r = 0.34$, $p < 0.0001$; SaHa: $r = 0.37$, $p < 0.0001$; DiHa: $r = 0.40$, $p < 0.0001$; SuHa: $r = 0.28$, $p = 0.004$; except for NeHa faces: $r = 0.12$, $p = 0.16$, *ns*), and all the expressions as a whole ($N = 1584$; $r = 0.35$ [0.16, 0.48], $p < 0.0001$).[1] In addition,

[1] As indicated in our "Supplemental Analyses and Tables" file (see Evidence scores of prototypical expressions: Specificity check), FACET clearly discriminated among seven *prototypical* facial expressions (neutral, happy, surprised, sad, fearful, disgusted, and

Table 1
Morphological action units. Mean evidence scores of action units (AUs; in Odds Ratios) for prototypical happy faces (rows) and blended expressions (columns), as computed by automated software analysis (Emotient FACET), and threshold levels of intensity (% of non-happy eyes) at which significant differences between happy and blended expressions emerged.

Blended expression							
Action units		AnHa	FeHa	SaHa	DiHa	SuHa	NeHa
AU1	Inner brow raiser	**70%**	**60%**	**60%**	ns	**50%**	ns
	Happy: −1.16	> −1.69	< 0.08	< −0.16		< 0.08	
AU2	Outer brow raiser	**50%**	**90%**	ns	**60%**	**50%**	ns
	Happy: −0.75	> −1.25	< 0.18		> −1.22	< 0.22	
AU4	Brow lowerer	**50%**	**50%**	**50%**	**50%**	ns	ns
	Happy: −1.70	< −0.72	< −0.97	< −0.91	< −0.96		
AU5	Upper lid raiser	**70%**	**60%**	ns	ns	**50%**	ns
	Happy: −1.36	< −0.75	< −0.57			< −0.51	
AU6	Cheek raiser	ns	**70%**	ns	ns	**50%**	ns
	Happy: 3.12		> 2.49			> 2.44	
AU7	Lid tightener	**90%**	ns	ns	**70%**	**40%**	ns
	Happy: 0.46	< 1.05			< 1.02	> 0.05	
AU9	Nose wrinkler	**50%**	ns	ns	**50%**	**70%**	ns
	Happy: −2.57	< −1.23			< −1.33	> −3.75	
AU12	Lip corner puller	**60%**	**70%**	**80%**	**80%**	**70%**	ns
	Happy: 4.29	> 3.62	> 3.57	> 3.62	> 3.72	> 3.61	
AU14	Dimpler	**100%**	**60%**	**80%**	ns	**50%**	ns
	Happy: −1.93	< −1.46	< −1.61	< −1.45		< −1.50	
AU23	Lip tightener	**60%**	ns	ns	ns	ns	ns
	Happy: −1.72	< −1.27					

Note. > or < Significantly higher or lower, respectively, AU scores for Happy than for Blended expressions (e.g., AU1 was significantly higher for prototypical happy faces than for AnHa faces at 70% intensity and above, etc.). ns: non-significant differences. Bold for threshold level of intensity.

Table 2
Relationships between measures. Pearson correlations ($N = 1584$ face stimuli) between (a) the probability of judging faces as happy (R "Happy"), (b) affective valence ratings (both a and b, Human Observers), (c) automated classification of faces as joy, and (d) action unit scores (both c and d, Emotient FACET Software).

		1 R "Happy" (Observers)	1 vs. 2	2 Joy evidence (Software)	2 vs. 3	3 Valence (Observers)
	Response "Happy	–				
	Joy Evidence	0.54***		–		
	Affective Valence	0.91***	0.0001	0.50***		–
AU1	Inner brow raiser	−0.19*	0.0001	−0.57***	0.0001	−0.05
AU2	Outer brow raiser	0.03	0.0001	−0.46***	0.0001	0.12
AU4	Brow lowerer	−0.69***		−0.62***		−0.63***
AU5	Upper lid raiser	−0.37**	0.0001	−0.65***	0.0001	−0.32**
AU6	Cheek raiser	0.17*	0.0001	0.72***	0.0001	0.14
AU7	Lid tightener	−0.16*	0.0001	0.03	0.0001	−0.22*
AU9	Nose wrinkler	−0.34**	0.0001	−0.05	0.0001	−0.38**
AU12	Lip corner puller	0.48***	0.0001	0.86***	0.0001	0.48***
AU14	Dimpler	−0.26*	0.0001	−0.53***	0.0001	−0.23*
AU23	Lip tightener	−0.22*	0.0001	−0.45***	0.0001	−0.29*

Note. 1 vs. 2 and 2 vs. 3 indicate significant differences (using the Fisher r-to-z transformation) between two correlation coefficients (1 vs. 2: R "Happy" column vs. Joy Evidence column; 2 vs. 3: Joy Evidence column vs. Valence column). For each single correlation, *$p < 0.01$; **$p < 0.001$; ***$p < 0.0001$.

affective valence ratings were correlated with human recognition of happiness, $r(1584) = 0.91$, $p < 0.0001$, CI [0.90, 0.92], and this occurred to a greater extent ($z = 27.50$, $p < 0.0001$) than with the automated joy evidence scores, $r(1584) = 0.50$, $p < 0.0001$, CI [0.46, 0.54]. In contrast, most of the AU measures correlated with joy evidence scores more than with human happiness judgments (both the probability of judging faces as happy and affective valence ratings; all Fisher's $zs > 7.00$, $p < 0.0001$; see Table 2).

Second, multiple regression analyses were performed to estimate the *contribution* of affective valence vs. morphological AUs to (a) human judgments of a smiling face as happy, relative to (b) the automated classification of a smiling face as joy. To this end, affective valence and the 10 selected AU scores were stepwise entered as predictors in two regression analyses. In one analysis, the probability of *responding "happy"* by *human observers* was the dependent variable. The model that produced the best regression fit, R^2 adj. = 0.864, $F(3, 1580) = 3347.94$, $p < 0.0001$, combined three predictors accounting for 75% of the variance in the dependent variable: The specific contributions (according to the sr^2 statistic) of (a) valence (49.56%; positive effect), (b) AU1 (3.53%; negative effect), and (c) AU9 (1.42%; negative effect), were significant (all $ps < 0.0001$). In the other analysis, *joy* evidence scores obtained by *automated assessment* served as the dependent variable. The model that produced the best regression fit, R^2 adj. = 0.843, $F(3, 1580) = 2824.36$, $p < 0.0001$, combined three predictors accounting for 71% of the dependent variable: The specific contributions (sr^2) of (a) AU12 (25.91%; positive effect), (b) AU2 (10.76%; negative effect), and (c) AU9 (5.52%; negative effect), were significant (all $ps < 0.0001$).

(*footnote continued*)
angry), with significant effects (all $ps < 0.0001$). Further, for such expressions, the correlation between human observers' and FACET categorization performance was $r(168) = 0.61$, $p < 0.0001$, $N = 168$ (24 KDEF models by 7 expressions). This reveals FACET sensitivity and convergence with human observers. In contrast, for *blended* expressions, we observed lower sensitivities and correlations between FACET and human observers. This may not be surprising, however, because—apart from being ambiguous—many blended expressions included low expressive eye intensities, thus increasing discrimination difficulties.

4. Discussion

This study investigated (a) discrimination thresholds for smiles in prototypical happy faces (with happy eyes) vs. smiles in blended expressions (with non-happy eyes); and (b) smile discrimination sensitivity differences between human perceivers and automated assessment with FACET computer software. Results revealed that (a) smile discrimination was interactively affected by intensity and type of eye expression, with some blends having lower discrimination thresholds than others; and (b) discrimination was better for human perceivers than for FACET assessment, with the former relying on affective valence, and the latter relying on morphological facial features.

4.1. Discrimination thresholds as a function of type of eye expression in human observers

Prior research has compared faces with a smiling mouth and happy-looking (Duchenne) eyes relative to smiling faces with *neutral* eyes, and has shown that observers can to some extent differentiate between the two types of smiles (Krumhuber et al., 2014; Okubo et al., 2012; Quadflieg et al., 2013; for a review, see Gunnery & Ruben, 2016). In the current study, this approach was extended by means of varying the *intensity* and *type* of eye expression, from happiness to neutrality or to each of five basic emotions (anger, sadness, disgust, fear, and surprise). For human observers, smiling faces with *angry* eyes (AnHa) were discriminated from prototypical happy faces at lower intensities than those with *sad* (SaHa), *fearful* (FeHa), or *disgusted* (DiHa) eyes, which were discriminated at lower intensities than those with *surprised* eyes (SuHa), which were discriminated at lower intensities than those with *neutral* eyes (NeHa). The reaction time asymptotes were in correspondence with this discrimination threshold pattern.

These findings complement those of prior research in which the non-happy eye intensity of smiling faces was held always at the apex: With different paradigms, (a) the probability of *confusing* such blended expressions as “happy” (Calvo et al., 2012; Calvo, Gutiérrez-García, et al., 2013), (b) the positive *affective priming* (Calvo et al., 2012), and (c) the minimum face *display time* required for discrimination (Gutiérrez-García & Calvo, 2015), were *directly* related to the current study thresholds, while (d) the probability of *first fixation* on the eye region (Calvo, Gutiérrez-García, et al., 2013), and (e) the amplitude of emotional expression processing *ERP components* (P200 and EPN; Calvo, Marrero, & Beltrán, 2013), were *inversely* related. That is, consistently, expressions with the (current) lowest threshold (AnHa) had less (a), (b) and (c), but more (d) and (e), whereas the opposite occurred for expressions with the highest threshold (NeHa), with the other blended expressions in between.

This reliable pattern across various paradigms and measures raises the issue of which factors facilitate or impair smile discrimination depending on the eye expression. In the current study, *affective valence* was the best predictor of the probability that human observers judged faces as happy. Discrimination threshold differences across types and intensities of blended expressions varied as a function of the affect they conveyed (with increasing discrimination difficulties as positive affect increased, and vice versa). This suggests that affective processing significantly contributes to discriminate between types of smiling faces (Calvo et al., 2018).[2] Such discrimination differences are not due to mere physical factors, as (a) the role of affective valence as a predictor remained significant when morphological facial features (AUs) were controlled in the multiple regression analyses; and (b) there was no interaction between type of eye expression and intensity for any of multiple low-level (e.g., luminance, etc.) and visual saliency properties of the face stimuli. At a more general level, the current findings fit into the broader literature regarding facial emotion perception and how the brain processes emotion. Evidence from two different approaches, i.e., affective priming paradigms (e.g., Lipp, Price, & Tellegen, 2009; McLellan et al., 2010) and emotional ERP components of brain activity (e.g., Calvo & Beltrán, 2013; Willis, Palermo, Burke, Atkinson, & McArthur, 2010), suggests that affective valence is extracted from facial expressions—including smiling faces—early and automatically (for a review, see Calvo & Nummenmaa, 2016). Our own findings converge in showing that affective processing is important for fine discrimination between types of smiling faces.

4.2. Human versus software discrimination of smiles: processing mechanisms

Automated analysis of *prototypical* (happy or non-happy) expressions with FACET software showed satisfactory classification and discrimination specificity (see Supplemental Analyses and Tables). The evidence scores of each expression was significantly greater for the corresponding stimulus category than for others (e.g., a sad face stimulus was more likely to be classified as “sad” than as any other expression, etc.). In addition, AUs characterized expressive categories in accordance with FACS proposals (Ekman et al., 2002). Further, for *blended* expressions, the automated assessment of joy was sensitive to eye expression intensity, and revealed an interaction with type of eye expression, just as it was the case for human observers. All this and the significant correlation between automated analysis and observer processing of joy/happiness demonstrate acceptable expression recognition performance using computer software (Bartlett & Whitehill, 2011; Cohn & De la Torre, 2015; Dailey, Cottrell, Padgett, & Adolphs, 2002; Olderbak et al., 2014; Susskind, Littlewort, Bartlett, Movellan, & Anderson, 2007).

Nevertheless, for blended expressions, human perceivers differentiated between types of smiles better than computer software did. The relationships between non-happy eye intensity and the joy evidence scores obtained by automated assessment, albeit significant, were lower than those for human observers. Also, relative to human achievement, FACET assessment yielded higher discrimination thresholds (hence poorer performance). Further, FACET classified most of the blended expressions as “less happy” than the prototypical happy expressions (except for smiling faces with neutral eyes) at some intensity level, but blends were never categorized as “not happy”. This suggests that activation of the distinctive AU12 (lip corner puller) characterizing the smiling mouth biases the FACET categorization of expressions as joy. In fact, in the regression analyses, AU12 was the best predictor of FACET joy evidence scores (accounting for 25.91% of the variance), whereas AU6 (cheek raiser)—albeit correlated with joy scores in the bivariate analyses—did not emerge as a significant predictor. Importantly, AU6 (eye region) is *different* for each type of smiling face, and therefore it should facilitate discrimination. In contrast, AU12 (mouth region) is *common* to both types of smiles, and therefore it would hinder discrimination. The highly significant role of AU12 (which probably overshadowed that of AU6) could thus explain the limited automated discrimination performance.[3]

What are the differences in the discrimination mechanisms between human and automated assessment of smiles? The correlation and

[2] Future research could also examine whether cue-congruency (not just valence) between the upper and the lower face half might better explain discrimination threshold differences as a function of type and intensity of eye expression.

[3] AU12 is the hallmark of a smile and is typically located in the lower half of a face. However, AU12 evidence scores varied with manipulations in the upper face half (see Table 1), despite the lower face half having been held constant across all the expressions with a smile. In fact, the low-level image properties (i.e., luminance, etc.) of the lower half of all the smiling faces were practically identical (all $ps \geq 0.98$). This implies that the modifications in the upper face region (different types and intensities of eye expressions) were used in some manner by the FACET “black box” algorithm to adjust the calculation for AU12 evidence. Further, this suggests that, although FACET computations probably rely mainly on local morphological features for each AU, some kind of configural processing of the face as a whole is made.

multiple regression analyses suggest that human processing is more dependent on affective valence of the face *as a whole*, whereas automated processing is more dependent on *single* facial *features*. This *analytic* coding can also be inferred from the strong—and probably biasing—contribution of AU12. Automated feature analysis would facilitate the classification of prototypical happy faces (with a smile *and* happy eyes) because the distinctive smile would gain prominence in the computation process; but it would compromise the correct rejection of blended expressions (a smile *but* non-happy eyes) as "not happy" because of the reduced weight of the eyes (in favor of the smiling mouth). In contrast, human observers, albeit relying also on single features (in fact, among the various AUs, AU12 showed the highest correlation with happiness judgments), depend on *holistic* processing to obtain an affective impression from the face (Calvo et al., 2012; Tanaka et al., 2012). Holistic or configural integration would allow for a better detection of inconsistencies between the (non-happy) eyes and the (smiling) mouth, thus facilitating the refined discriminations (including affective processing) required for blended expressions.

4.3. Limitations and extensions

An alternative explanation of the human observers' discrimination advantage over FACET for blended vs. prototypical happy expression can be considered. The majority of the face stimuli (i.e., blended and prototypical happy faces) involved the *same* smiling mouth (although also prototypical non-happy faces without a smile in the lower face half were presented), while they had a *variety* of eye expressions. This gave observers ample opportunity to learn that the upper face was especially task-relevant whereas the lower face had less discriminative value. As a consequence, observers might have adopted an attentional strategy involving preferential attention towards the upper face, thus facilitating discrimination. In contrast, FACET presumably does not focus on particular facial areas, but rather considers all of them for categorization. Eye-tracking assessment could be useful to examine this explanation in human observers, and FACET could be ideally trained to adopt "attentional strategies" depending on type and intensity of expressive changes. As another extension, it would be interesting to know if FACET might perform better with regards to discriminating blended expressions when AU12 is presented at a lower intensity. This would be useful to test our hypothesis of an overrepresentation of the smile in the FACET computations. Relatedly, further refinement of FACET algorithms could re-adjust the weights assigned to different facial features of happy faces, or to enhance the configural integration of the mouth with the eyes, and to train the system with blended expressions.

Beyond basic emotions, a person can feel mixed emotions (Heavey, Lefforge, Lapping-Carr, & Hurlburt, 2017; Russell, 2017; Watson & Stanton, 2017), and brain systems can in fact support simultaneous and mixed affective processing and experience (Man, Nohlen, Melo, & Cunningham, 2017). Further, some blended emotions combine happiness with negative emotions (e.g., schadenfreude, nostalgia, enjoying being frightened—e.g., by a horror film or a roller-coaster—etc.), which may give way to very different types of smiles. In addition, social norms often constrain the magnitude of emotional exhibition and, therefore, expressive changes in the face may be subtle, which adds difficulty to the interpretation of the underlying emotions. The issue of emotion blends and expression subtlety is important regarding smile discrimination, given the multifaceted and multifunctional nature of smiles (Crivelli et al., 2015; Niedenthal et al., 2010). In the current study, we addressed this issue by varying the type and intensity of the eye region expression in faces with a smile. The scope could be extended by varying also the shape and intensity of the smiling mouth itself, to encompass a greater diversity of mixed emotions and expressions.

Finally, there is the issue of what the comparison of automated and human discrimination performance adds to our knowledge of facial emotion processing. Computational models of facial expression processing, such as FACET and others (e.g., EMPATH: Dailey et al., 2002; SVM: Susskind et al., 2007; or CERT: Littlewort et al., 2011) classify facial expressions as certain emotions based on specific *perceptual* features and configurations (see Bartlett & Whitehill, 2011; Cottrell & Hsiao, 2011). Accordingly, we can argue that facial expressions can be coded by "emotionless machines", in the absence of affective processing. Yet we found a significant correlation between automated classification and subjective *affective* valence ($r = 0.50$; and also between automated and human categorization performance: $r = 0.54$). This suggests that human observers also rely greatly on perceptual processing of morphological features. It is true, however, that the correlation between human discrimination performance and affective ratings was significantly higher ($r = 0.91$), which reveals an additional significant role of emotional processing, beyond—even if based on—perceptual processing.

5. Conclusions

Smiles in blended expressions with non-happy eyes can be discriminated by human observers as *not* conveying happiness, although this greatly depends on type and intensity of expression in the eye region. The discrimination threshold is lowest (i.e., easier discrimination, at lower intensities) for angry eyes, and it is highest (i.e., more difficult) for neutral eyes, with fearful, sad, disgusted, and surprised eyes in between. Discrimination of blended expressions is more accurate for human perceivers than for automated assessment using FACET software, which, however, showed significant accuracy and specificity in classifying prototypical expressions. Human performance is highly related to perception of affective valence, whereas automated performance is highly related to morphological action units (AUs). This suggests that the former relies more on configural face processing, whereas the latter relies more on analytical feature processing.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.actpsy.2018.04.019.

Funding

This research was supported by Grant PSI2014-54720-P to MGC from the Spanish Ministerio de Economía y Competitividad, and Grant RE 3721/2-1 to GR from the Deutsche Forschungsgemeinschaft.

References

Ambadar, Z., Cohn, J. F., & Reed, L. I. (2009). All smiles are not created equal: Morphology and timing of smiles perceived as amused, polite, and embarrassed/nervous. *Journal of Nonverbal Behavior, 33*, 17–34.

Bartlett, M., & Whitehill, J. (2011). Automated facial expression measurement: Recent applications to basic research in human behavior, learning, and education. In A. Calder, G. Rhodes, M. Johnson, & J. Haxby (Eds.). *Handbook of face perception* (pp. 489–513). Oxford, UK: Oxford University Press.

Borji, A., & Itti, L. (2013). State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 35*, 185–207.

Calvo, M. G., & Beltrán, D. (2013). Recognition advantage of happy faces: Tracing the neurocognitive processes. *Neuropsychologia, 51*, 2051–2060.

Calvo, M. G., & Fernández-Martín, A. (2013). Can the eyes reveal a person's emotions? Biasing role of the mouth expression. *Motivation and Emotion, 37*, 202–211.

Calvo, M. G., Fernández-Martín, A., & Nummenmaa, L. (2012). Perceptual, categorical, and affective processing of ambiguous smiling facial expressions. *Cognition, 125*, 373–393.

Calvo, M. G., Gutiérrez-García, A., Avero, P., & Lundqvist, D. (2013). Attentional mechanisms in judging genuine and fake smiles: Eye-movement patterns. *Emotion, 13*, 792–802.

Calvo, M. G., Gutiérrez-García, A., & Del Líbano, M. (2018). What makes a smiling face look happy? Visual saliency, distinctiveness, and affect. *Psychological Research, 82*, 296–309.

Calvo, M. G., Marrero, H., & Beltrán, D. (2013). When does the brain distinguish between genuine and ambiguous smiles? An ERP study. *Brain and Cognition, 81*, 237–246.

Calvo, M. G., & Nummenmaa, L. (2008). Detection of emotional faces: Salient physical features guide effective visual search. *Journal of Experimental Psychology: General, 137*, 471–494.

Calvo, M. G., & Nummenmaa, L. (2016). Perceptual and affective mechanisms in facial expression recognition: An integrative review. *Cognition and Emotion, 30*, 1081–1106.

Cohn, J. F., Ambadar, Z., & Ekman, P. (2007). Observer-based measurement of facial expression with the Facial Action Coding System. In J. A. Coan, & J. J. B. Allen (Eds.). *The handbook of emotion elicitation and assessment* (pp. 203–221). New York: Oxford University Press.
Cohn, J. F., & De la Torre, F. (2015). Automated face analysis for affective computing. In R. A. Calvo, S. DiMello, J. Gratch, & A. Kappas (Eds.). *The Oxford handbook of affective computing* (pp. 131–150). New York: Oxford University Press.
Cohn, J. F., & Schmidt, K. L. (2004). The timing of facial motion in posed and spontaneous smiles. *International Journal of Wavelets, Multiresolution and Information Processing, 2*, 1–12.
Cottrell, G. W., & Hsiao, J. H. (2011). Neurocomputational models of face processing. In A. Calder, G. Rhodes, M. Johnson, & J. Haxby (Eds.). *The Oxford handbook of face perception* (pp. 401–425). Oxford, UK: Oxford University Press.
Crivelli, C., Carrera, P., & Fernández-Dols, J. M. (2015). Are smiles a sign of happiness? Spontaneous expressions of judo winners. *Evolution and Human Behavior, 36*, 52–58.
Dailey, M. N., Cottrell, G. W., Padgett, C., & Adolphs, R. (2002). EMPATH: A neural network that categorizes facial expressions. *Journal of Cognitive Neuroscience, 14*, 1158–1173.
Ekman, P. (2001). *Telling lies: Clues to deceit in the marketplace, politics, and marriage.* New York: W.W. Norton & Co.
Ekman, P., & Friesen, W. V. (1978). *Facial action coding system.* Palo Alto, CA: Consulting Psychologists Press.
Ekman, P., Friesen, W. V., & Hager, J. C. (2002). *Facial action coding system. Investigator's guide.* Salt Lake City, UT: Human Face.
Fernández-Dols, J. M., & Crivelli, C. (2013). Emotion and expression: Naturalistic studies. *Emotion Review, 5*, 24–29.
Frank, M. G., Ekman, P., & Friesen, W. V. (1993). Behavioral markers and recognizability of the smile of enjoyment. *Journal of Personality and Social Psychology, 64*, 83–93.
Girard, J. M., Cohn, J. F., Jeni, L. A., Sayette, M. A., & De la Torre, F. (2015). Spontaneous facial expression in unscripted social interactions can be measured automatically. *Behavior Research Methods, 47*, 1136–1147.
Gordon, I., Tanaka, J., Pierce, M., & Bartlett, M. (2011). Facial expression production and training. *Journal of Vision, 11*, 565. http://dx.doi.org/10.1167/11.11.565.
Gunnery, S. D., & Hall, J. A. (2014). The Duchenne smile and persuasion. *Journal of Nonverbal Behavior, 38*, 181–194.
Gunnery, S. D., & Ruben, M. A. (2016). Perceptions of Duchenne and non-Duchenne smiles: A meta-analysis. *Cognition and Emotion, 30*, 501–515.
Gutiérrez-García, A., & Calvo, M. G. (2015). Discrimination thresholds for smiles in genuine versus blended facial expressions. *Cogent Psychology, 2*, 1064586.
Hassin, R. R., Aviezer, H., & Bentin, S. (2013). Inherently ambiguous: Facial expressions of emotions in context. *Emotion Review, 5*, 60–65.
Heavey, C. L., Lefforge, N. L., Lapping-Carr, L., & Hurlburt, R. T. (2017). Mixed emotions: Toward a phenomenology of blended and multiple feelings. *Emotion Review, 9*, 105–110.
Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*, 1489–1506.
Krumhuber, E. G., Likowski, K. U., & Weyers, P. (2014). Facial mimicry of spontaneous and deliberate Duchenne and non-Duchenne smiles. *Journal of Nonverbal Behavior, 38*, 1–11.
Krumhuber, E. G., & Manstead, A. S. R. (2009). Can Duchenne smiles be feigned? New evidence on felt and false smiles. *Emotion, 9*, 807–820.
Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual.* Technical Report A-8Gainesville, FL: University of Florida.
Lipp, O. V., Price, S. M., & Tellegen, C. L. (2009). No effect of inversion on attentional and affective processing of facial expressions. *Emotion, 9*, 248–259.
Littlewort, G., Whitehill, J., Wu, T., Fasel, I., Frank, M., Movellan, J., & Bartlett, M. (2011). The computer expression recognition toolbox (CERT). *Proceedings of the IEEE international conference on automatic face and gesture recognition*California: Santa Barbara (Doi: 10.1109/FG.2011.5771414).
Lundqvist, D., Flykt, A., & Öhman, A. (1998). *The Karolinska directed emotional faces – KDEF.* Stockholm, Sweden: CD-ROM from Department of Clinical Neuroscience, Psychology Section, Karolinska Institutet (ISBN 91-630-7164-9).
Man, V., Nohlen, H. U., Melo, H., & Cunningham, V. A. (2017). Hierarchical brain systems support multiple representations of valence and mixed affect. *Emotion Review, 9*, 124–132.
McLellan, T. L., Johnston, L., Dalrymple-Alford, J., & Porter, R. (2010). Sensitivity to genuine vs. posed emotion specified in facial displays. *Cognition and Emotion, 24*, 1277–1292.
McLellan, T. L., Wilcke, J. C., Johnston, L., Watts, R., & Miles, L. K. (2012). Sensitivity to posed and genuine displays of happiness and sadness: A fMRI study. *Neuroscience Letters, 531*, 149–154.
Miles, L. K., & Johnston, L. (2007). Detecting happiness: Perceiver sensitivity to enjoyment and non-enjoyment smiles. *Journal of Nonverbal Behavior, 31*, 259–275.
Niedenthal, P. M., Mermillod, M., Maringer, M., & Hess, U. (2010). The simulation of smiles (SIMS) model: Embodied simulation and the meaning of facial expression. *Behavioral and Brain Sciences, 33*, 417–433.
Okubo, M., Kobayashi, A., & Ishikawa, K. (2012). A fake smile thwarts cheater detection. *Journal of Nonverbal Behavior, 36*, 217–225.
Olderbak, S., Hildebrandt, A., Pinkpank, T., Sommer, W., & Wilhelm, O. (2014). Psychometric challenges and proposed solutions when scoring facial emotion expression codes. *Behavior Research Methods, 46*, 992–1006.
Quadflieg, S., Vermeulen, N., & Rossion, B. (2013). Differential reliance on the Duchenne marker during smile evaluations and person judgments. *Journal of Nonverbal Behavior, 37*, 69–77.
Russell, J. A. (2017). Mixed emotions viewed from the psychological constructionist perspective. *Emotion Review, 9*, 111–117.
Schneider, W., Eschmann, A., & Zuccolotto, A. (2002). *E-Prime user's guide.* Pittsburgh, PA: Psychology Software Tools.
Slessor, G., Bailey, P. E., Rendell, P. G., Ruffman, T., Henry, J. D., & Miles, L. K. (2014). Examining the time course of young and older adults' mimicry of enjoyment and non-enjoyment smiles. *Emotion, 14*, 532–544.
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*, 152–162.
Tanaka, J. W., Kaiser, M., Butler, S., & Le Grand, R. (2012). Mixed emotions: Holistic and analytic perception of facial expressions. *Cognition and Emotion, 26*, 961–977.
Underwood, G., & Foulsham, T. (2006). Visual saliency and semantic incongruency influence eye movements when inspecting pictures. *The Quarterly Journal of Experimental Psychology, 59*, 1931–1949.
Valstar, M. F., Pantic, M., Ambadar, Z., & Cohn, J. F. (2006). Spontaneous vs. posed facial behavior: Automatic analysis of brow actions. *Proceedings of the ACM international conference on multimodal interfaces, Banff, Canada*.
Walther, D., & Koch, C. (2006). Modelling attention to salient proto-objects. *Neural Networks, 19*, 1395–1407.
Watson, D., & Stanton, K. (2017). Emotion blends and mixed emotions in the hierarchical structure of affect. *Emotion Review, 9*, 99–104.
Willis, M. L., Palermo, R., Burke, D., Atkinson, C. M., & McArthur, G. (2010). Switching associations between facial identity and emotional expression: A behavioural and ERP study. *NeuroImage, 50*, 329–339.