

Optimal faces for gender and expression: A new technique for measuring dynamic templates used in face perception

Frédéric J.A.M. Poirier

Visual Psychophysics and Perception Laboratory, School
of Optometry, Montréal, Québec, Canada



Jocelyn Faubert

Visual Psychophysics and Perception Laboratory, School
of Optometry, Montréal, Québec, Canada



Facial expressions are important for human communications. Face perception studies often measure the impact of major degradation (e.g., noise, inversion, short presentations, masking, alterations) on natural expression recognition performance. Here, we introduce a novel face perception technique using rich and undegraded stimuli. Participants modified faces to create optimal representations of given expressions. Using sliders, participants adjusted 53 face components (including 37 dynamic) including head, eye, eyebrows, mouth, and nose shape and position. Data was collected from six participants and 10 conditions (six emotions + pain + gender + neutral). Some expressions had unique features (e.g., frown for anger, upward-curved mouth for happiness), whereas others had shared features (e.g., open eyes and mouth for surprise and fear). Happiness was different from other emotions. Surprise was different from other emotions except fear. Weighted sum morphing provides acceptable stimuli for gender-neutral and dynamic stimuli. Many features were correlated, including (1) head size with internal feature sizes as related to gender, (2) internal feature scaling, and (3) eyebrow height and eye openness as related to surprise and fear. These findings demonstrate the method's validity for measuring the optimal facial expressions, which we argue is a more direct measure of their internal representations.

Keywords: face perception, facial expression, gender, emotions, dynamic template

Citation: Poirier, F. J. A. M., & Faubert, J. (2012). Optimal faces for gender and expression: A new technique for measuring dynamic templates used in face perception. *Journal of Vision*, 12(6):28, 1–28, <http://www.journalofvision.org/content/12/6/28>, doi:10.1167/12.6.28.

Introduction

Despite much work on face perception, we still don't know “computationally and neurally, what might a face ‘template’ look like (...)?” (McKone, Kanwisher, & Duchaine, 2007). Here, we present a novel method to measure face templates, which we used to measure gender templates and dynamic emotion templates. Analyses on these templates also provide clues on what features might be useful for recognition, categorization, and holistic processing of gender and emotions.

Faces convey a lot of information, especially when in motion, about identity, gender, and emotion. However, face stimuli are very complex, making face perception a difficult area of research. The difficulty is compounded by the ease in which we can extract information from face stimuli. Researchers often have to weigh many conflicting trade-offs when making stimuli, e.g., stimuli that appear natural, that are well-controlled, and that avoid ceiling effects. Also, as with any multidimensional stimuli, face perception researchers need to trade-off the number of exemplars and dimensions

varied within a study with its duration. Some of these common trade-offs are reviewed briefly here, then our new method is introduced which bypasses many of these trade-offs.

The first trade-off is between information control and naturalistic stimuli. Much of the work on the recognition of facial expressions have taken photos or movies of facial expressions and then modified them to equate, mask, or otherwise modify the information contained. The impact on recognition of these modifications is then measured to provide hints on our internal representation of faces. Natural stimuli are difficult to control for information content. This is especially relevant for face perception, where simple cues such as skin texture, skin color, or haircut can indicate gender or identity. The concern also extends to task-irrelevant cues that might be learned when using small subsets of stimuli, as is often the case in reverse correlation studies with natural faces. Also, unless photos are taken in the laboratory, faces often do not fall neatly into frontal, 3/4, or side view, and may also contain large differences in lighting conditions. Even when information can be controlled, natural stimuli

require extensive manipulation before being applicable for research: external features are generally cropped out, facial hair is removed either before or after the picture is taken, skin defects need to be corrected, and faces have to be aligned properly. To avoid these issues, many researchers have turned to synthetic faces (e.g., Brunswik & Reiter, 1937; Garrod, Yu, Breidt, Curio, & Schyns, 2010; Goren & Wilson, 2006; Kostov, Yanagisawa, Johansson, & Fukuda, 2001; Sigala & Logothetis, 2002; Wilson, Loffler, & Wilkinson, 2002).

The second trade-off is between performance control and the extent of stimulus modifications. With unmodified faces, participants usually reach ceiling performance in face perception tasks even at short presentation times. For this reason, stimuli are often degraded such that performance hovers near 50–75% correct (for review, see Roark, Barrett, Spence, Hervé, & O’Toole, 2003; see also Loffler, Gordon, Wilkinson, Goren, & Wilson, 2005, for masking). These methods have greatly contributed to our understanding of natural facial expression recognition by identifying the natural components that most strongly influence performance. However, these methods also suffer from their limitations. In degraded-stimuli conditions, participants often focus on the most informative parts and thus may ignore other parts of the stimuli that otherwise contribute to the effect. Other methods for degrading performance can also modify the nature of the processing involved (e.g., inversion effect). To avoid ceiling effects, the implemented procedures generally increase difficulty well beyond that found naturally by using short presentation times and degraded stimuli. It is unclear to what extent faster degraded face perception extends to face perception in normal viewing conditions.

The third trade-off is between the dimensionality of the face space and the time required of participants to complete studies. For example, we know very little about what the optimal facial expression might be. Optimal facial expressions may differ significantly from natural faces by exaggerating useful components, i.e., the caricature advantage (Benson & Perrett, 1994; Lee, Byatt, & Rhodes, 2000; Mauro & Kubovy, 1992; Rhodes, Brennan, & Carey, 1987; Rhodes & Tremewan, 1994, 1996; for review, see Rhodes, 1996; Valentine, 2001; Valentine, Darling, & Donnelly, 2004). Searching the face space for optimal stimuli using standard methods would be a daunting task. Masking methods are not meant to modify the information per se, only its availability. Classification image techniques offer a way toward optimal stimuli, unfortunately at a great cost in testing time for anything beyond the main components of an expression. Also, classification image techniques generally rely on image modification rather than morphing and

as such produce results that cannot easily be extrapolated (e.g., Mangini & Biederman, 2004).

The fourth trade-off is between the use of dynamic stimuli and its increased dimensionality and presentation time. Emotional expression is dynamic. Most facial expressions studies use static photos (e.g., Ekman & Friesen, 1976). Research results suggest that dynamic and static emotional expressions are not processed similarly (e.g., Ambadar, Schooler, & Cohn, 2005; Biele & Grabowska, 2006; Harwood, Hall, & Shinkfield, 1999; O’Toole, Roark, & Abdi, 2002; for review, see Roark et al., 2003), thus supporting the study of dynamic facial expressions for greater ecological validity. Indeed, dynamic emotions are recognized more easily than static emotions (Ambadar et al., 2005; Cunningham & Wallraven, 2009a; Harwood et al., 1999; Kätsyri & Sams, 2008) even when the static information contained in the dynamic displays is degraded (Cunningham & Wallraven, 2009b; Ehrlich et al., 2000). Facial motion can also be used to recognize identity (Knappmeyer, Thornton, & Bühlhoff, 2003; Lander, 2005). Moreover, using point-light displays, studies have shown that the temporal pattern of facial motion is sufficient to support the recognition of facial emotions (Bassili, 1978, 1979). However, studies often use morphs (e.g., LaBar, Crupain, Voyvodic, & McCarthy, 2003; Sato, Kochiyama, Yoshikawa, Naito, & Matsumura, 2004) or subsets of emotions (e.g., Kilts, Egan, Gideon, Ely, & Hoffman, 2003), and it is unclear to what extent their findings generalize to natural dynamics and to other facial expressions respectively. This has prompted some researchers to create a larger bank of validated dynamic facial expressions (e.g., Roy et al., 2007).

The purpose of the present study is to assess the merits of a novel approach, where participants manipulate the facial expression stimulus rather than perform some decision task based on it. Participants are asked to create a face to represent a given gender or emotion as best as they can, using 53 facial features they can manipulate. Thus, participants iteratively adjust the stimulus to create the best match to their template. The method aims to measure participant’s templates under nondegraded conditions. The 53 chosen dimensions exceed the dimensionality of comparable methods (e.g., Wilson et al. [2002] used 37 dimensions) and can replicate most geometrical facial changes of the Facial Actions Coding System (Ekman & Friesen, 1978; see [Methods](#) for details).

The chosen emotions include the six basic emotions (Ekman & Friesen, 1975; Izard, 1971), which are argued to be similar across backgrounds and cultures (Ekman & Friesen, 1975; Izard, 1971, 1994), and are the basis for the Facial Actions Coding System (Ekman & Friesen, 1978). To this set, we also added pain, gender, and neutral.

Several analyses are then performed on the data collected on these templates and categories, to emphasize: (1) the static and dynamic features present in these categories, as well as their usefulness for categorization, (2) the clustering of categories along several statistically derived dimensions as a measure of similarity, (3) the distances and correlations between categories in multidimensional space to separate intensity from similarity, (4) the covariation between features to identify those that are linked and thus could benefit from integrative processes, and (5) the applicability of morphing techniques to create dynamic expressions or neutral categories.

Methods

Participants

Six participants volunteered (four males and two females), including the first author as well as university undergraduate and graduate students. Their vision was normal or corrected to normal.

Apparatus

Testing and data collection were conducted on a PC computer (Pentium D 3GHz) set to a resolution of 1280 × 960 pixels and a refresh rate of 60Hz. Responses were recorded via mouse button presses. Viewing distance was 56 cm, at which distance 32 pixels corresponded to 1° of visual angle. The dynamic face area subtended 512 × 512 pixels, thus 16° high by 16° wide. The entire stimulus area subtended 32° wide by 16° high in dynamic conditions, or 24° by 16° in static conditions.

Procedure

Facial adjustment task

On each trial, participants were shown a display containing five components: (1) a static or dynamic face showing an animation of the face from the starting to ending state, (2 and 3) two static faces showing the starting and ending state respectively (not shown in static conditions), (4) text labels indicating the characteristics of the face that participants needed to create, and (5) up to 10 bars showing the name and value(s) of dimensions that could be adjusted (Figure 1). Participants were instructed to adjust the values of the dimensions in order to most convincingly recreate the characteristics given by the labels.

Stimuli

Stimuli consisted of five parts: (1) the static or dynamic face, (2 and 3) the starting and ending state of the faces, (4) the text labels, and (5) the dimensions.

The characteristics of the static or dynamic face could be adjusted along 53 dimensions (Table 1). The dimensions chosen are geometric due to hardware limitations. Our faces are untextured and two-dimensional (much like Wilson's), which makes it difficult to include features such as crinkled noses (important for disgust), reflectance properties (important for gender), etc. We hope that, if the current method gains acceptance and interest, a future version of this method could include these features.

Head shape was adjusted along 16 radial dimensions in all directions starting at 0° in steps of 22.5° and labeled with cardinal directions (e.g., N, ESE, SW). The radial dimensions controlling head shape were subdivided into two categories, seven describing the head shape for the top part (above W-E orientations), and nine describing the head shape for the bottom part (below and including W-E orientations). Mouth shape was adjusted along 10 dimensions: (1 and 2) teeth average height and opening, (3 and 4) lips' average height and opening, (5 and 6) height of mouth corners relative to mouth height, (7 and 8) distances between midline and inner mouth on either side of midline, and (9 and 10) distances from mouth corners and inner mouth on either side of midline. Eyes were adjusted along nine dimensions: (1 and 2) height and distance from midline, (3 to 5) gaze direction (up/down, left/right) and alignment (convergence/divergence) relative to eye positions, (6) eyelid width, (7 and 8) eyelid opening on each eye independently, and (9) eyelid orientation. Eyebrows were adjusted along 10 dimensions: (1) thickness, (2) distance from midline to the central portion, (3 and 4) distance from central portion to the inner and outer portions, (5 and 6) height of the eyebrows' central portion, independently for each eyebrow, and (7 to 10) height of inner and outer portion relative to the eyebrow's central portion, independently for each eyebrow. Other physical characteristics were adjusted along eight dimensions: (1 and 2) upper and lower lip thickness, (3 to 6) nose width, height, length, and shape, and (7 and 8) pupil and iris width. Nose shape was adjusted by changing the height of the sides of the nose relative to the nose's center's height.

The dimensions of the faces prior to adjustments were set to match the dimensions of the average face with a neutral emotion, averaged across male and female faces taken from Wilson et al. (2002).

For dynamic conditions, the following 37 dimensions were allowed to have a different starting, intermediate, and ending state: (1 to 9) head shape

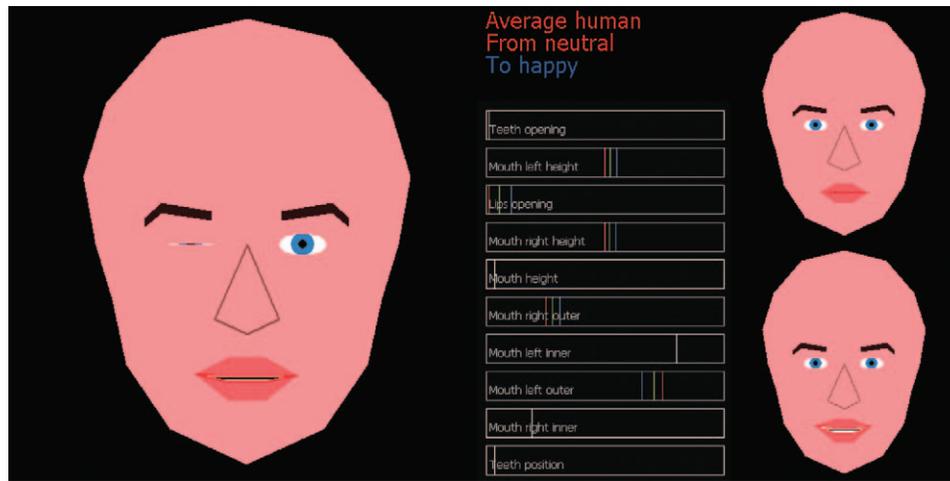


Figure 1. Sample stimulus as shown during the experiment. Stimuli consisted of five parts: (left) the static or dynamic face, (right) faces for the starting and ending states, (top middle) the text labels indicating the “goal” face, and (bottom middle) the dimensions along which participants could adjust the face. In the example, the left eye blinks and the mouth gradually changes from neutral to a smile. Here, participants had to create an average human face changing from neutral to happy. See text for details.

dimensions including and below midline (W-E), (10 and 11) nose height and shape, (12 and 13) iris and pupil width, (14 and 15) teeth position and opening, (16 to 19) left/right, inner/outer mouth width, (20 to 22) left/intermediate/right mouth height, (23) mouth opening, (24 to 26) gaze direction and alignment, (27 and 28) left/right eye opening, (29 to 34) left/right, outer/intermediate/inner eyebrow height, and (35 to 37) outer/intermediate/inner eyebrow separation. The following dimensions were always static: (1 to 7) head shape above the midline (W-E), (8 and 9) nose width and length, (10 to 13) eye width, angle, separation and

height, (14) eyebrow thickness, and (15 and 16) upper and lower lip thickness.

Colored surfaces were chosen instead of lines to more accurately represent the edge/line relationships present in natural images. Blue eyes (RGB = 0, 128, 255) were used to improve iris-pupil contrast. The skin was pink (RGB = 255, 150, 150) with a darker nose outline (RGB = 150, 75, 75). Lips were slightly darker and redder than the skin (RGB = 255, 100, 100) with the opening delineated by a darker line (RGB = 255, 50, 50). Eyebrows were brown (RGB = 50, 0, 0).

This face was shown at three positions in the screen for dynamic conditions (coordinates given from the

Category	Characteristic	Dimensions	Dynamic	Sample label
1	Face shape	7: ENE, NE, NNE, ... WNW	No	Face shape: N
2	Face shape	9: E, ESE, SE, SSE, ... W	Yes	Face shape: SE
3	Iris and pupil	2: pupil size, iris size	Yes	Iris: size
3	Nose	2: height, shape	Yes	Nose: shape
3	Nose	2: width, length	No	Nose: width
3	Lips	2: thickness (upper/lower)	No	Lips: thickness (upper)
4	Lips	1: opening	Yes	Lips: opening
4	Teeth	2: height, opening	Yes	Teeth: height
4	Mouth width	4: inner/outer left/right	Yes	Mouth width: outer left
4	Mouth height	3: left/right corner, average	Yes	Mouth height: left corner
5	Eyes	4: width, angle, separation, height	No	Eyes: width
5	Eyes	2: opening (right/left)	Yes	Eye: opening (right)
5	Gaze	3: left-right, low-high, alignment	Yes	Gaze (left-right)
6	Eyebrows	3: length (inner/outer), separation	Yes	Eyebrows: length (inner)
6	Eyebrows	1: thickness	No	Eyebrows: thickness
6	Eyebrow height	6: left/right inner/average/outer	Yes	Eyebrow height: left inner
	Totals	53 dimensions	37 dynamic + 16 static	

Table 1. List of dimensions used in the experiment, along with the category (1–6), whether the dimension was dynamic, and sample labels.

top-left corner), (1) the dynamic face was shown in a window spanning pixels 128 to 640 along the width, and pixels 144 to 656 along the height, (2 and 3) the static faces representing starting state and ending state were shown at half-scale in windows spanning pixels 896 to 1152 along the width and pixels 144 to 400 and 400 to 656 along the height, respectively. Static conditions did not include the starting and ending states in the display.

The response area included 7 to 10 response bars in a window spanning pixels 640 to 896 along the width, and pixels 144 to 656 along the height. Bars were 32 pixels high, and interspaced by eight pixels. When fewer than 10 response bars were shown, the extra bottom positions were left blank. Each bar included a grey text label, and up to three response lines indicated by different colors (i.e., red for starting state, green for intermediate state, and blue for ending state). The bar's outline was grey if the participant had made at least one adjustment on the relevant dimension during the trial; otherwise the outline was white. Each bar included a text label indicating which dimension it was linked to. The bars were shown in one of six categories (Table 1): (1) top seven head shape dimensions, (2) bottom nine head shape dimensions, (3) mouth shape, (4) eyes, (5) eyebrows, and (6) other physical characteristics. Participants could use keys 1 through 6 to change the category (with the categories randomly assigned to keys before each trial to control for order effects), and the dimensions included in the category (see Stimuli section for details) were shown in random order.

Above the response bars, text labels were given to indicate what the participants had to recreate. These text labels constitute the experimental conditions. For static conditions, the labels were “Average human, Neutral expression,” “Average female, Neutral expression,” and “Average male, Neutral expression.” For dynamic conditions, the labels all started with “Average human, From neutral” but could finish with one of the following alternatives: “To happy,” “To surprised,” “To fear,” “To angry,” “To sad,” “To disgusted,” and “To pain,” which are essentially the same labels as used by Roy et al. (2007).

Temporal sequence

The dynamic faces were animated with the following sequence: (1) 500 ms static in starting state, (2) 250 ms linearly morphing from the starting state to the intermediate state, (3) 250 ms linearly morphing from the intermediate state to the ending state, (4) 500 ms static in ending state, and (5) 500 ms of blank. This 2 sec animation was cycled continuously during the trial. Static faces were shown continuously without a blank period.

Participants were given as much time as required to complete a condition, and the condition was considered completed if they felt that they had optimized the stimulus. To encourage adjusting all dimensions, participants were told not to spend more than 30 seconds on a dimension, and that they could come back to it later if they felt it was needed. Also, untouched dimensions were marked by a brighter outline on the dimension bar to encourage participants to try adjusting every dimension at least once each per trial.

Analysis: normalization and comparisons of means

Normalization for head size was performed using the following steps: (1) the average radius was calculated across the 16 radii for each head, (2) the global average radius was computed across every head, e.g., across participants and conditions, and (3) each head was scaled so that its average radius was equal to the global average radius.

The following variables were created as combinations of other variables: (1) head size, (2) head width at temples, (3) head width at W-E height, (4) head length, (5) head aspect-ratio, (6) inner mouth width, (7) outer mouth width, (8) average lip thickness, (9) left eye-eyebrow distance, and (10) right eye-eyebrow distance.

Analyses on static data included only the ending state, whereas analysis on dynamic data was based on comparisons between states. Means of ending states and mean dynamics (ending state minus starting state) were analyzed using *t*-tests and ANOVAs, using individual variability as a measure of error. Effects were also considered nonsignificant if the average difference was much smaller than 1 pixel or if the difference occurred on teeth position or openness when the lips were closed.

Our experimental design includes several definitions of “neutral.” First, a “neutral” condition was included, defined as an emotionless gender-neutral face. Also, the “default” face is defined as the average of male and female faces from the Wilson et al. (2002) dataset, which were used as the starting values of the experiment. Comparison with the default face indicates which features participants changed. These results are shown in Figure 6 but are not discussed. In dynamic conditions, motion can be compared to a default of “no motion.” Finally, in the “neutral” section of the results, the neutral condition is compared to various morphed neutrals. To distinguish those, in that section, the neutral condition is referred to as the “measured neutral” and compared to average and optimal morphs named “average neutral” and “optimal neutral,” respectively (see the Neutral section of the Results section for details on morphing).

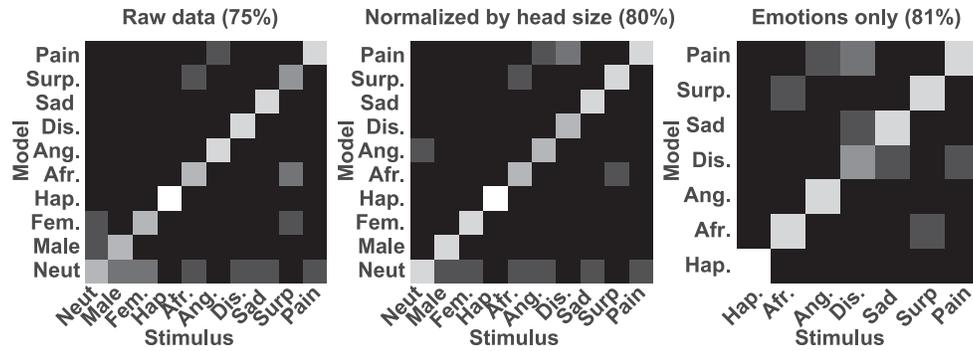


Figure 2. Classification performance of a statistical algorithm (discriminate analysis) based on features, both without or with normalization, and with or without neutral and gender conditions. Only one classification performance is shown for the emotions-only classification because classification rates with and without normalization were identical. Classification rates of individual faces created by participants are in the range of 75% to 81%. See text for details.

Analysis: discriminate analysis

Discriminate analysis (“classify” function in Matlab; using the “diagLinear” method) was used for classification. It generates a linear classifier for each pair of categories, maximizing the correct pair-wise categorization. These classifiers find features that would be most useful for performance in forced-choice tasks. The presence of useful features does not necessarily imply that participants use these features; in fact, it is possible that participants ignore some useful features, or conversely that participants sometimes use misleading or nonpredictive features. Because exemplars across participants were pooled in this analysis, we needed a different measure of variability for statistical analysis. To analyze classification weights, *t*-tests and ANOVAs used variability across conditions; that is, we analyzed how reliable weights were to discriminate one condition (e.g., pain) from all other included conditions (e.g., all other emotions). As such, a significant weight indicates that the feature is consistently associated with the face category and consistently dissociated with other face categories included in the analysis. Classification was performed on static (ending state) and dynamic (ending state minus starting state) data normalized for head size, then *z*-scored. Classification was either performed on all 10 categories (static and dynamic ending state included) or just the seven emotions (static or dynamic). The classification performance is shown in Figures 2 and 10. We used $\alpha = 0.001$ for statistical significance, which corresponds to Bonferroni correction for family-wise error against 53 comparisons. We used $\alpha = 0.005$ for near-significance.

Analysis: dynamic trajectory

Analyses were performed to detect significant nonlinearities in the dynamic change in facial expressions.

Such nonlinearities would provide evidence that linear morphing techniques are insufficient to represent the dynamical aspects of the emotions. In particular, we measured the detour length, i.e., the physical deviation from a linear path. None of the average detour lengths differed from 0 for the dynamic conditions ($ts(5) < 3.08$, $ps > 0.054$), and detour lengths were similar across emotions ($F(6, 30) = 0.56$, $p = 0.76$). Other angle-based measures were also nonsignificant. That is, there was no nonlinearity detected in the dynamic trajectories. These analyses are not discussed further.

Analysis: distances and correlations

To estimate similarity between faces, two measures were used: distance and correlation. Distance refers to the geometric difference between two faces. As the distance between two faces becomes smaller, they become more similar and as such should be harder to discriminate from each other.

Correlation uses the neutral face as origin and gives an index of how two faces deviate similarly (or dissimilarly) from the neutral face. This calculation is similar conceptually to measuring the angle formed by two vectors, each originating at neutral and pointing toward the two different faces. Correlation measures the similarity of the vectors for the two faces independent of intensity. A correlation significantly above 0 means that the two faces are similarly different from the neutral, although they can still differ in intensity, e.g., related emotions such as contentment and happiness would be expected to have a high positive correlation even though they differ in intensity. A correlation near 0 means that the two faces are orthogonal, i.e., they differ from neutral in different ways. A correlation significantly below 0 means that the two faces are opposites, e.g., we would expect a

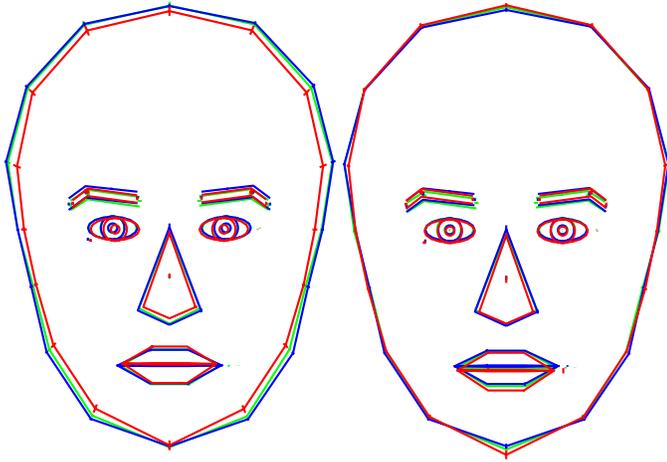


Figure 3. Gender and face characteristics, shown un-normalized (left) and normalized for head size (right). Un-normalized, females faces (red) are smaller, thinner, have thinner eyebrows and noses, and have a more elongated aspect-ratio than males (blue). Normalized, female faces are more elongated and have thinner eyebrows and noses. Error bars show the standard error of the mean. See text for details.

negative correlation for emotions that produce facial deformations in opposite directions.

Individual participant data was first normalized to a common coordinate frame before averaging. Common coordinate frames were used in these calculations to ensure that the standard error reflects variability in only the relevant dimensions. They also prevent effects to be added if they're not oriented in the same direction, e.g., distances should sum to the extent that they are aligned. Note that normalizing to a common coordinate frame has an effect on the measures of variability, but not on the means themselves.

The common coordinate frame for distance was the axis from one mean static face to the other. The common coordinate frame for correlation was the plane formed by two axes, one from the neutral face to the first emotion, and the other from the neutral face to the second emotion (adjusted to be orthogonal to the first axis). Thus, every correlation used in the average and standard deviation was measured using vectors that are expressed in a common coordinate frame.

With 10 conditions, there was a total of 45 distances ($10 \times 9/2$) and a total of 36 correlations (excluding neutral, which was used as reference: $9 \times 8/2$). ANOVAs were used to determine if correlations and distances were equal to each other, and a regression analysis was then used to summarize the distances and correlations data in a compact way. The regression analyses used one variable per condition to indicate whether the distance or correlation included that condition (10 for distances, 9 for correlations), which were entered as a block. Forward entry was then used to account for interactions in which all combinations of two condi-

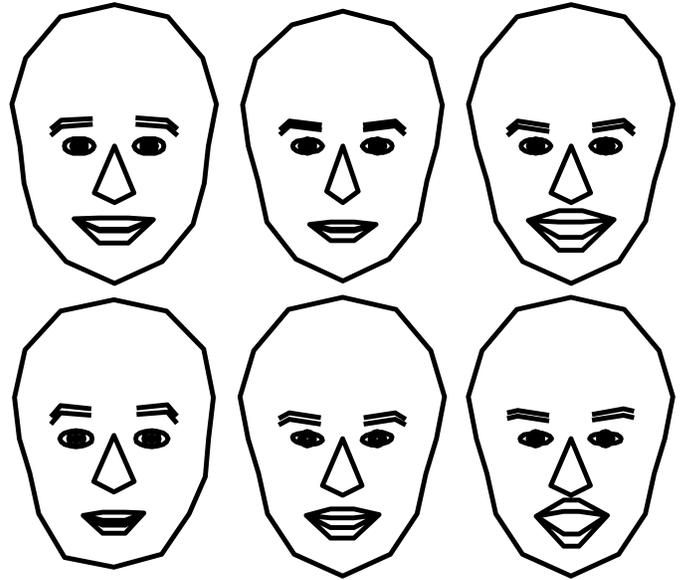


Figure 4. Faces for the happy condition for all participants, ending state shown. Although every face can be recognized as happy, there are many differences such as the intensity of the emotion, the face shape, and the eyes orientation.

tions were entered separately (45 for distances, 36 for correlations) and only kept if statistically significant at the 0.05 level. The interaction terms detect and adjust for nonadditive combinations, i.e., when the distance or correlation between two faces is significantly different from that predicted based on the average distance or correlation of those two faces with all other faces. The cutoff for higher and lower distances and correlations is chosen by visual inspection of the histogram.

Analysis: factor analysis

Features were also analyzed by how they correlate together. Common factor analysis (“factoran” function in Matlab) was used to cluster features together that are correlated. This is commonly used as a data-reduction technique, but here the emphasis was to describe how facial features covary, either by moving together during emotions or by being somehow related. Thus, it was used as a way to reduce the dimensionality of the stimulus into meaningful dimensions. Factor analysis used data across participants and conditions, and was analyzed using standard factor analysis methods. Two solutions were included, one with 21 factors and another with six. As factor analysis results are known to sometimes depend on the number of factors allowed, comparing the 21-factor and the six-factor analyses provides a way to evaluate the importance and independence of factors. The dataset had a rank of 53 out of 53 dimensions, thus a 21-factor and a six-factor solution represent data reductions of 60.4% and

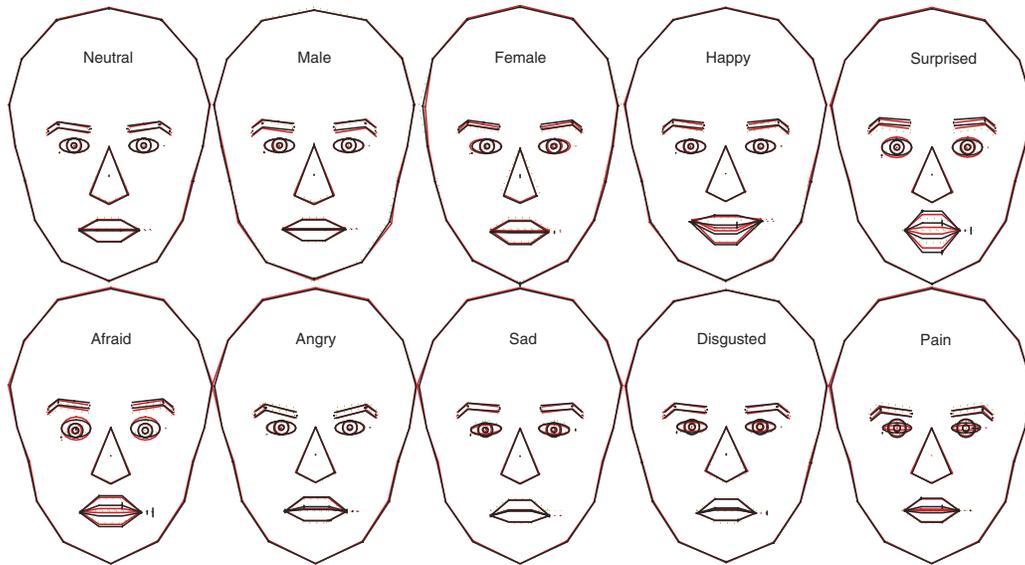


Figure 5. Mean face characteristics (black) and classifier results (red) compared to the default face (green) for gender and emotions. Results shown are for ending states normalized for head size. Male and female faces are best classified based on face shape, as well as nose and eyebrow characteristics. Much of the emotional expression is carried by the eyebrows, the eyes' openness, and the mouth. Some features are uniquely associated with given conditions, e.g., (1) wider mouth curved upward with happiness, (2) higher eyebrows and open jaw (teeth) with surprise, (3) lower inner eyebrows with anger, (4) sloping eyes with sadness, (5) nose shape with disgust, and (6) lower eyebrows with pain. Some features are shared with several emotions, e.g.: (1) open lips with surprise, fear, and pain, (2) higher eye-eyebrow distance with surprise and fear, (3) wide open eyes with surprise and fear, (4) partly shut eyes with disgust and pain, and (5) mouth curved downward with sad and disgust. See text for details.

88.7%, respectively. Only weights greater than 0.2 were discussed.

Results

Participants made systematic adjustments to many facial features for all conditions, compared to the default face, except for the neutral condition for which no feature was significantly changed ($ts(5) < 2.2$, $ps > 0.08$).

Static data are presented first, including ending state emotions. Gender (see Gender section) and emotions (see Emotional faces (static) section) were compared to each other using multidimensional similarity measures (see Distances and correlations section). Finally, static data was analyzed to recover commonalities across conditions in the way facial position information changes (see Factor analysis section). Dynamic information was analyzed using the differences between starting and ending states (see Emotional faces (dynamic) section). Finally, different definitions of the neutral were compared and contrasted for static and dynamic faces (see Neutral section).

Terminology is briefly reviewed in the relevant sections, and details of analyses are left in the relevant methods sections.

Linear discriminate analysis was used to extract features useful for classification, and those features were reported in “compared to all faces” (see Gender and Emotional faces [static] sections) and “compared to all emotions” analyses (see Emotional faces (dynamic) section). The classification performance of the linear discrimination analysis on static conditions was at least 75% correct. Normalization increased performance from 75% correct to 80% correct. Most of the misclassification errors were made when faces belonging to non-neutral categories were classified as neutral. There is also some confusion between surprise and fear, and between sad, anger, disgust, and pain. Removing the neutral and gender conditions improved performance to 81% equally for normalized and raw data. That is, linear discrimination analysis could classify individual static faces produced by participants with 75% to 81% accuracy.

Gender

Changes to static-only conditions (gender and neutral) are reviewed here, and dynamic conditions are reviewed in the following text (starting at the Emotional faces (static) section). Each static face was analyzed compared to all faces, using linear discriminate analysis, to extract unique features. A separate

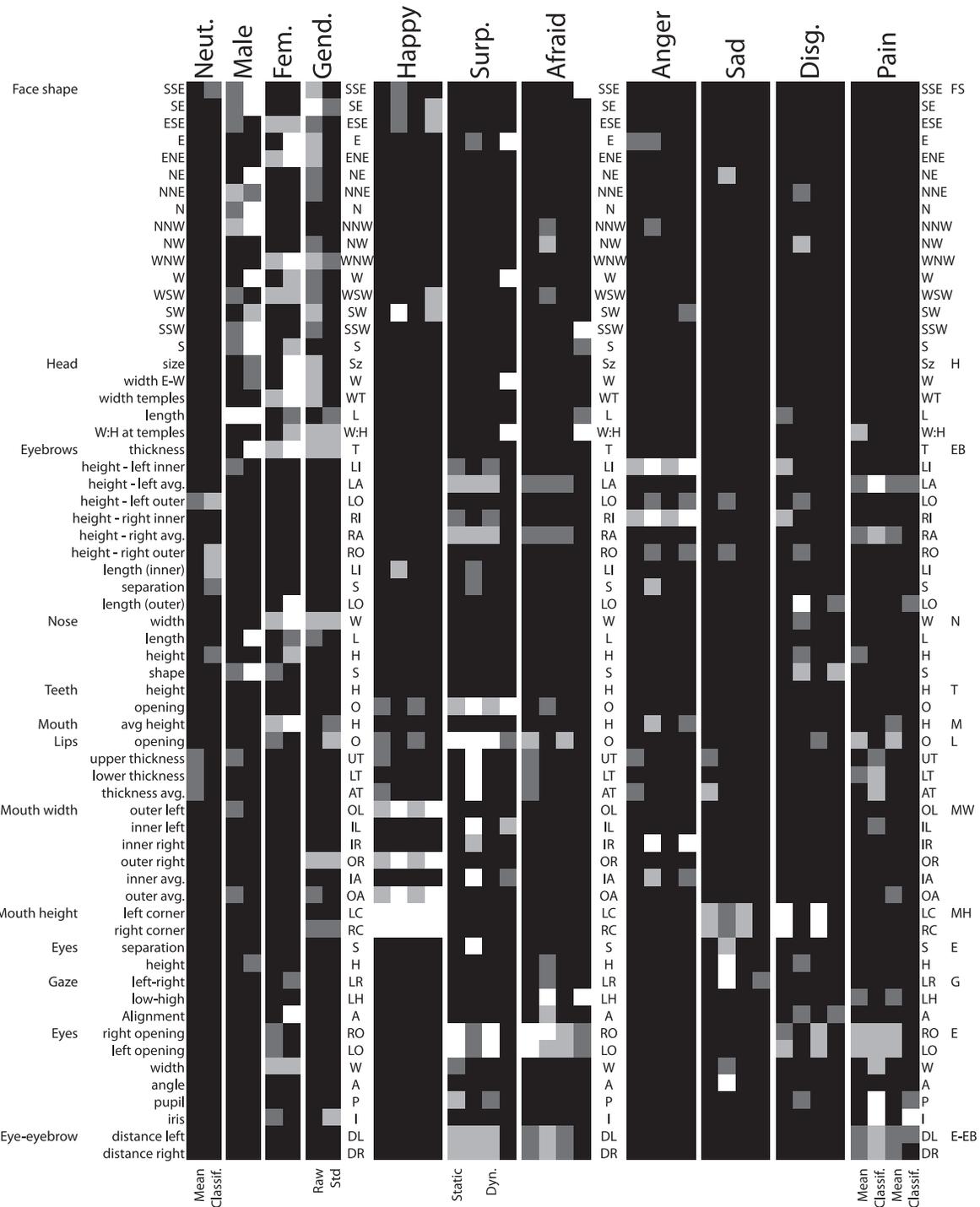


Figure 6. Overview of statistical analyses. Rows show the 53 dimensions used, plus 10 dimensions that are combinations of dimensions (i.e., head size, head width at W-E and temples, head length, head aspect-ratio, average lip thickness, inner and outer average mouth width, right and left eye-eyebrow distance). The 11 main columns are for the 10 conditions plus one for the comparison across gender. Static faces include two tests each: (1) mean vs. default and (2) classifier (i.e., linear discriminate analysis weights). Gender includes two tests: (1) on raw data and (2) on data normalized for head size. Emotional faces include four tests each: (1) static mean vs. default, (2) static classifier, (3) dynamic mean vs. no motion, and (4) dynamic classifier. Statistical significance level is shown using four intensities: black for not statistically significant (>0.1 and >0.005), dark grey for near-significant (>0.05 and >0.001), light grey for significant (<0.05 and <0.001), and white for highly significant (<0.01 and <0.0001 , respectively, for analyses on means and on classifier data). Generally, a result is interesting if it replicates across analyses (e.g., happy shows an increase in mouth width and upward movement of mouth corners in all four analyses). Some discrepancies are informative, e.g., disgust and sadness are associated with the lowering of mouth corners, but that shared feature is not useful for classification). See text for details.

analysis “comparing to the default face” produced similar results.

Neutral face characteristics included lower-outer eyebrows ($t(8) = 6.2$, $ps < 0.0003$ for left and right), and shorter eyebrow length for the inner part ($t(8) = 6.9$, $p = 0.0001$). No other facial feature was reliably associated with the neutral condition.

According to our participants, males and females differed on many facial features for the *raw data* (i.e., not normalized for head size). Female faces had slightly smaller heads than males ($t(5) = 3.06$, $p < 0.028$), which possibly influenced the position and size of other facial features. Gender also affected head width ($ts(5) > 2.66$, $ps < 0.045$ at temples and W-E), aspect-ratio ($t(5) = 3.36$, $p < 0.021$), eyebrow thickness ($t(5) = 3.78$, $p < 0.013$), and nose width ($t(5) = 3.83$, $p < 0.013$). Thus, it is reasonable to ask how males and females differ at comparable head sizes. Except for factor analysis, which included factor(s) for head size, the following analyses were conducted on data normalized for head size. (See Analysis: normalization and comparisons of means section for normalization procedure.)

Male face characteristics included a lower head top ($ts(8) = 7.6$ and 11.5 , $ps < 0.0001$ for NNW and N, respectively; NNE not significant, $t(8) = 4.8$, $p = 0.0014$), a wider chin ($ts(8) > 6.7$, $ps < 0.0001$ for SW, SSW, SSE, and SE), a shorter chin ($t(8) = 7.8$, $p < 0.0001$), a shorter head ($t(8) = 10.1$, $p < 0.0001$), a longer nose ($t(8) = 13.8$, $p < 0.0001$) with lower nostrils ($t(8) = 8.3$, $p < 0.0001$), and thicker eyebrows ($t(8) = 10.3$, $p < 0.0001$). Near-significance by our criteria included a lower head top ($t(8) = 4.8$, $p = 0.0014$ for NNE), lower eyes ($t(8) = 3.87$, $p = 0.0047$), and a bigger head ($t(8) = 4.56$, $p = 0.0019$).

Female face characteristics included a more oval head shape with thinner sides ($ts(8) > 5.5$, $ps < 0.0006$ for W, WNW, E, and ENE; widths at temples and W-E: $ts(8) > 7.3$, $ps < 0.0001$), larger cheeks ($ts(8) > 5.8$, $ps < 0.0004$ for WSW and ESE), and a longer head (aspect-ratio: $t(8) = 6.9$, $p = 0.0001$; $ts(8) > 3.5$, $ps < 0.008$ for N and S), as well as thinner eyebrows ($t(8) = 9.6$, $p < 0.0001$) whose outside length is shorter ($t(8) = 9.4$, $p < 0.0001$), a thinner and lower nose ($ts(8) < 6.2$, $ps < 0.0003$), a lower mouth ($t(8) = 9.25$, $p < 0.0001$), wider eyes ($t(8) = 6.58$, $p < 0.0002$), and gaze alignment slightly shifted inward ($t(8) = 8.05$, $p < 0.0001$).

There were significant gender-related differences on data normalized for head size. Comparing males and females, females' heads had a more elongated aspect-ratio (measured at the temples; $t(5) = 3.36$, $p = 0.020$), although head length was near-significant ($t(5) = 2.4$, $p = 0.063$), and head width was not significant ($ts(5) < 1.77$, $ps > 0.138$). Female's highbrows were thinner ($t(5) = 3.4$, $p = 0.018$). Females had a thinner nose ($t(5) = 3.07$, $p = 0.028$). Otherwise, female faces were similar to male faces ($ts(5) < 1.6$, $ps > 0.17$). Males and

females did not differ on physical characteristics related to eyes ($ts(5) < 2.0$, $ps > 0.11$) or mouth ($ts(5) < 2.4$, $ps > 0.07$) except for the outer-right mouth corner being further from the face center in males ($t(5) = 2.8$, $p = 0.039$).

Emotional faces (static)

Static emotions refer to the ending state of the emotion. Participants were instructed to generate a dynamic change in the faces so that they start at neutral and change to the target emotion, thus the ending state is also the emotional expression itself. Each static face was analyzed compared to all faces, using linear discriminate analysis, to extract unique features.

Happy face characteristics included increased inner eyebrow length ($t(8) = 5.1$, $p = .0009$), widening and upward curving of mouth corners ($ts(8) > 14$, $ps < 0.0001$), and the SW face part moves further ($t(8) = 7.3$, $p < 0.0001$). Eyebrows could not be used reliably for classification ($ts(8) < 2.1$, $ps > 0.079$).

Surprise face characteristics included increased average eyebrow height ($ts(8) = 5.98$ and 5.91 , $ps < 0.0004$), open lips and teeth ($ts(8) = 9.3$ and 13.3 , $ps < 0.0001$), thicker lips ($ts(8) = 11.3$, 13.9 , and 12.8 , $ps < 0.0001$, respectively, for upper, lower, and average), wider inner mouth ($ts(8) = 10.5$, 5.4 , and 8.4 , $ps < 0.0001$, < 0.0007 and < 0.0001 , respectively, for left, right, and average, respectively), eyes further apart ($t(8) = 9.2$, $p < 0.0001$), and larger eye-eyebrow distance ($ts(8) = 5.1$ and 5.2 , $ps < 0.0009$). Near-significance by our criteria included increased eyebrow separation ($t(8) = 4.8$, $p < 0.0013$) and increased inner eyebrow length ($t(8) = 4.3$, $p < 0.0026$), open eyes ($ts(8) = 4.59$ and 4.29 , $ps < 0.0018$ and < 0.0027).

Fear face characteristics included wide open eyes ($ts(8) > 7.0$, $ps < 0.0002$), lower gaze ($t(8) = 23.6$, $p < 0.0001$), and uncrossed gaze ($t(8) = 5.5$, $p = 0.0006$), higher eye-eyebrow distance ($ts(8) = 5.4$ and 5.3 , $ps < 0.0007$), and shorter NW face-shape distance ($t(8) = 6.8$, $p < 0.0001$). The following components were not significant by our criterion: higher eyebrow height ($ts(8) = 4.0$, $ps = 0.004$ for left and right sides), and lips opening ($t(8) = 3.7$, $p = 0.0057$).

Angry face characteristics included lower inner eyebrows ($ts(8) = 16.3$ and 16.7 , $ps < 0.0001$ for left and right, respectively), shorter distance between eyebrows ($t(8) = 5.7$, $p < 0.0004$), higher mouth ($t(8) = 6.5$, $p < 0.0002$), shorter inner mouth ($ts(8) = 6.5$ and 12.6 , $ps < 0.0002$ and < 0.0001 for average and left, respectively; right was not significant: $t(8) = 2.1$, $p = 0.068$). Near-significance by our criteria included lower outer eyebrows ($ts(8) = 3.96$ and 4.38 , $ps = 0.0041$ and 0.0024 , respectively, for left and right).

Sad face characteristics included eyes closer together ($t(8) = 5.3, p < 0.0007$), lower eyes ($t(8) = 8.5, p < 0.0001$), eyes angled with inner corners upward ($t(8) = 27, p < 0.0001$), NE face shape decrease ($t(8) < 6.4, p = 0.0002$), and gaze shift toward the left ($t(8) = 7.8, p < 0.0001$). Near-significance by our criteria included higher outer-right eyebrow ($ts(8) = 4.6$ & $4.7, ps < 0.0018$ and < 0.0016), lower mouth corners ($ts(8) = 4.4$ and $4.2, ps < 0.0022$ and < 0.0027 for left and right, respectively), and wider eyes ($t(8) = 4.2, p < 0.003$).

Disgusted face characteristics included only higher nostrils (i.e., nose shape, $t(8) = 5.9, p < 0.0004$) and NW face shape increase ($t(8) = 6.1, p < 0.0003$). Near-significance by our criteria included outer eyebrow height ($ts(8) = 5.0$ and $4.0, ps = 0.0010$ and 0.0035 for left and right, respectively), nose height ($t(8) = 4.9, p = 0.0013$), gaze convergence ($t(8) = 4.9, p = 0.0011$), and pupil size increase ($t(8) = 4.8, p = 0.0014$).

Pain face characteristics included lower average eyebrows ($ts(8) = 7.4$ and $6.8, ps < 0.0002$), thinner lips ($ts(8) > 5.0, p \leq 0.001$), eyes partly shut ($ts(8) = 6.3, ps < 0.0002$), eyes less wide ($t(8) = 5.6, p < 0.0005$), wider iris and pupil ($ts(8) = 7.3$ and $30.7, ps < 0.0001$), and lower eye-eyebrow distances ($ts(8) = 5.8$ and $5.4, ps = 0.0004$ and 0.0006 for left and right, respectively).

Distances and correlations

Figure 7 shows two measures used to assess similarity: distances and correlations between static faces. The distance between two faces is the Euclidian sum of geometric changes (in pixels) required to morph one face into the other. The correlation between two faces is calculated as the correlation between the two neutral-to-face vectors. Variability of distance and correlation measurements is calculated in a common coordinate frame (see Analysis: distances and correlations section for details).

Some faces are closer to each other than others ($F(44, 220) = 3.67, p < 0.0001$), with mean distances ranging from 11.7 to 40.6. The 12 highest distances were found between surprise and everything except fear (eight distances), and between happy and the four negative emotions (four distances). More specifically, the 12 highest distances were surprise-neutral (39.2), surprise-male (40.1), surprise-female (38.7), surprise-happy (38.1), surprise-anger (40.6), surprise-sad (40.0), surprise-disgust (36.5), surprise-pain (38.2), happy-anger (38.3), happy-sad (40.6), happy-disgust (38.9), and happy-pain (36.0). The 13 lowest distances were found between neutral and gender (two distances), neutral and the four negative emotions (four distances), the four negative emotions between each other (six distances), and between surprise and fear. More specifically, the 13 lowest distances were neutral-male

(11.7), neutral-female (15.8), neutral-anger (16.7), neutral-sad (17.7), neutral-disgust (16.9), neutral-pain (19.8), disgust-male (20.8), surprise-fear (18.1), anger-sad (18.7), anger-disgust (17.2), anger-pain (18.4), sad-disgust (13.7), sad-pain (20.4), and disgust-pain (17.9).

Distance between faces is well-predicted by a constant for each face ($R^2 = 0.971, R^2_{adj} = 0.962, F(10, 35) = 116, p < 0.001$), with happy and surprise being further from other faces ($As = 22.9$ and 24.2), fear being moderately far ($A = 13.7$), disgust being somewhat closer to other faces ($A = 9.5$), and neutral being closest to other faces ($A = 7.7$), all other faces being roughly equidistant from each other (As vary from 10.8 and 12.6). The main face pair that was ill-fitted by these main effects was surprise and fear, which were much closer to each other than expected by their distance to other faces. The inclusion of a weight for surprise and fear significantly improved the fit ($A = -25.6; R^2 = 0.985, R^2_{adj} = 0.980; F_{inc}(1, 34) = 31.1, p_{inc} < 0.001$). Other significant nonadditivities could be corrected by adding specific weights, further significantly improving the quality of the fit ($R^2 = 0.990$ to $0.996; R^2_{adj} = 0.986$ to $0.994, F_{incS} = 15.1$ to $5.0, p_{incS} = 0.001$ to 0.034). These additional weights were for face pairs that were also closer to each other than predicted by the main effects: happy-surprise ($A = -18.4$), happy-fear ($A = -12.5$), happy-male ($A = -9.7$), male-neutral ($A = -7.3$), happy-female ($A = -6.9$), and sad-disgust ($A = -5.7$). Under this full model which includes a weight per emotion plus seven weights for the specific face pairs previously discussed, happy and surprise were still further away from other faces ($As = 27.6$ and 28.8), with fear also far away ($A = 17.6$), and the neutral and disgust were still closer to other faces ($As = 7.3$ and 9.0), all other faces being roughly equidistant from each other (As vary from 10.1 and 12.2). The added seven weights specific to face-pairs did not significantly change the main effects.

Mean correlations varied between conditions ($F(35, 175) = 2.307, p < 0.001$) but remained fairly low (mean = 0.1749 ; range = -0.2360 to 0.6514), meaning that most faces differ from each other in some qualitative way rather than just in intensity. Correlations between faces are well-predicted by a constant for each face ($R^2 = 0.538, R^2_{adj} = 0.384, F(9, 27) = 3.496, p = 0.006$). No specific correlation deviated significantly from the previous prediction, meaning that some faces were more unique, while others were more similar to many faces. The highest eight correlation pairs are surprise-fear (0.6514), pain-disgust (0.6279), pain-anger (0.4063), pain-sad (0.4813), sad-disgust (0.6149), male-fear (0.5050), male-happy (0.4193), and happy-surprise (0.5153). The five lowest correlation pairs are happy-sad (-0.2128), happy-anger (-0.1799), happy-disgust (-0.1601), anger-male (-0.2360), and anger-female (-0.1498).

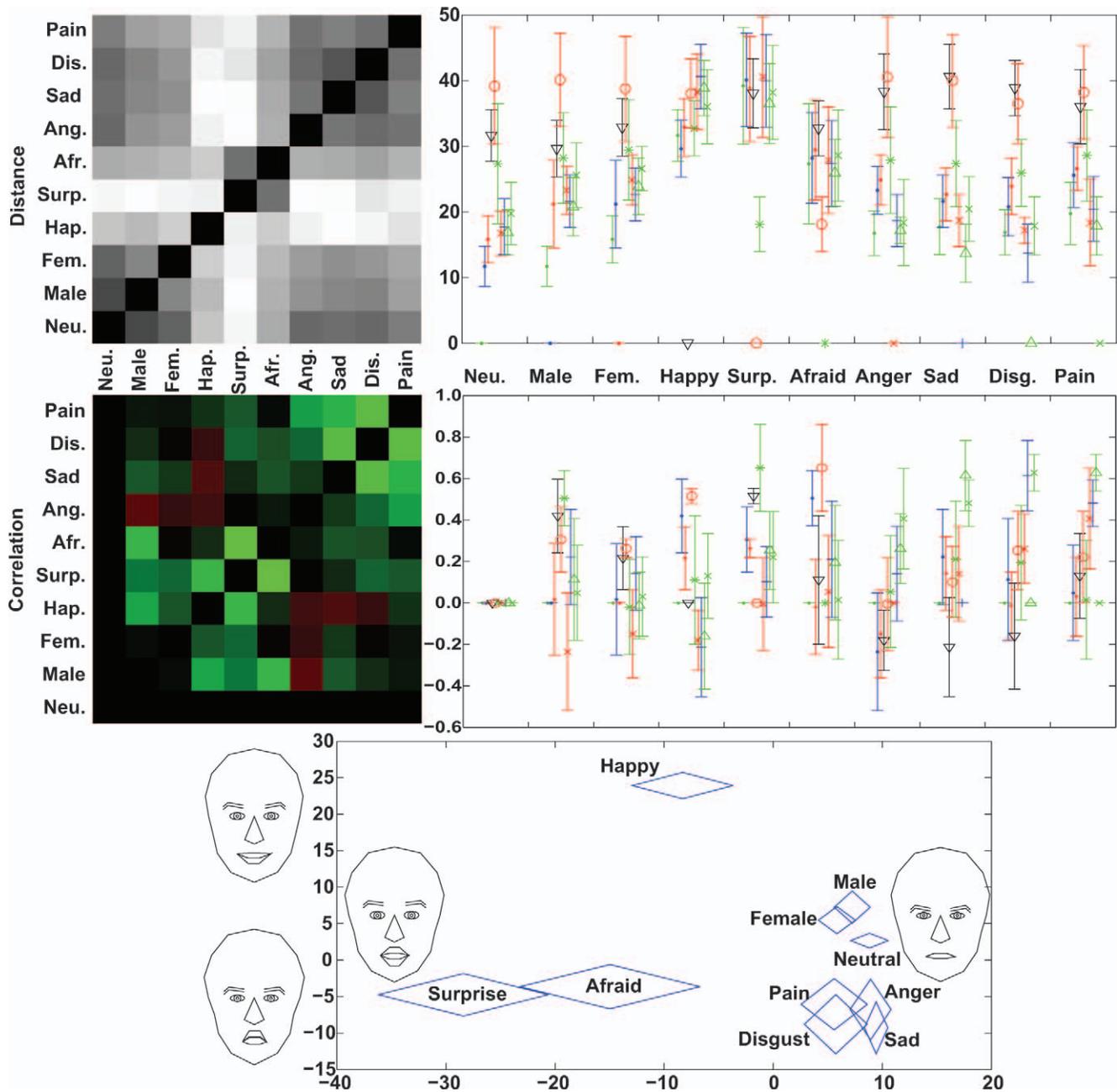


Figure 7. Distance in pixels (top) and correlation (intermediate) between each pair of the 10 static faces. Correlation between two faces is calculated using the neutral face as reference. (bottom) The 10 faces plotted on the two first components of a principal components analysis (in pixels), shown to clarify the results at the top. It is clear that happy is a unique face. Surprise is also unique except for fear being a close neighbor. All other faces have many close neighbors. High correlations are formed between fear and surprise, between happy and surprise, and between disgust, sad, and pain. High correlations suggest that these emotions should be more difficult to discriminate if shown at equally low intensities given their similarity. Happy is negatively correlated with anger, sad, and disgust. Most other face-pairs are in some way near-orthogonal to each other, meaning that they differ in some qualitative way rather than just in intensity. See text for details.

The first two axes of a principal component analysis provide an overview to help understand distance and correlation measurement effects. The first axis is mainly related to expansion (or contraction) of the face: mouth and eyes open (or close) and eyebrows move up (or

down). The second axis is mainly related to mouth corners extending upward and outward (or inward and downward). Faces plotted along these two axes cluster into four groups: (1) happy alone, (2) surprise and fear, (3) gender and neutral, and (4) the other four negative

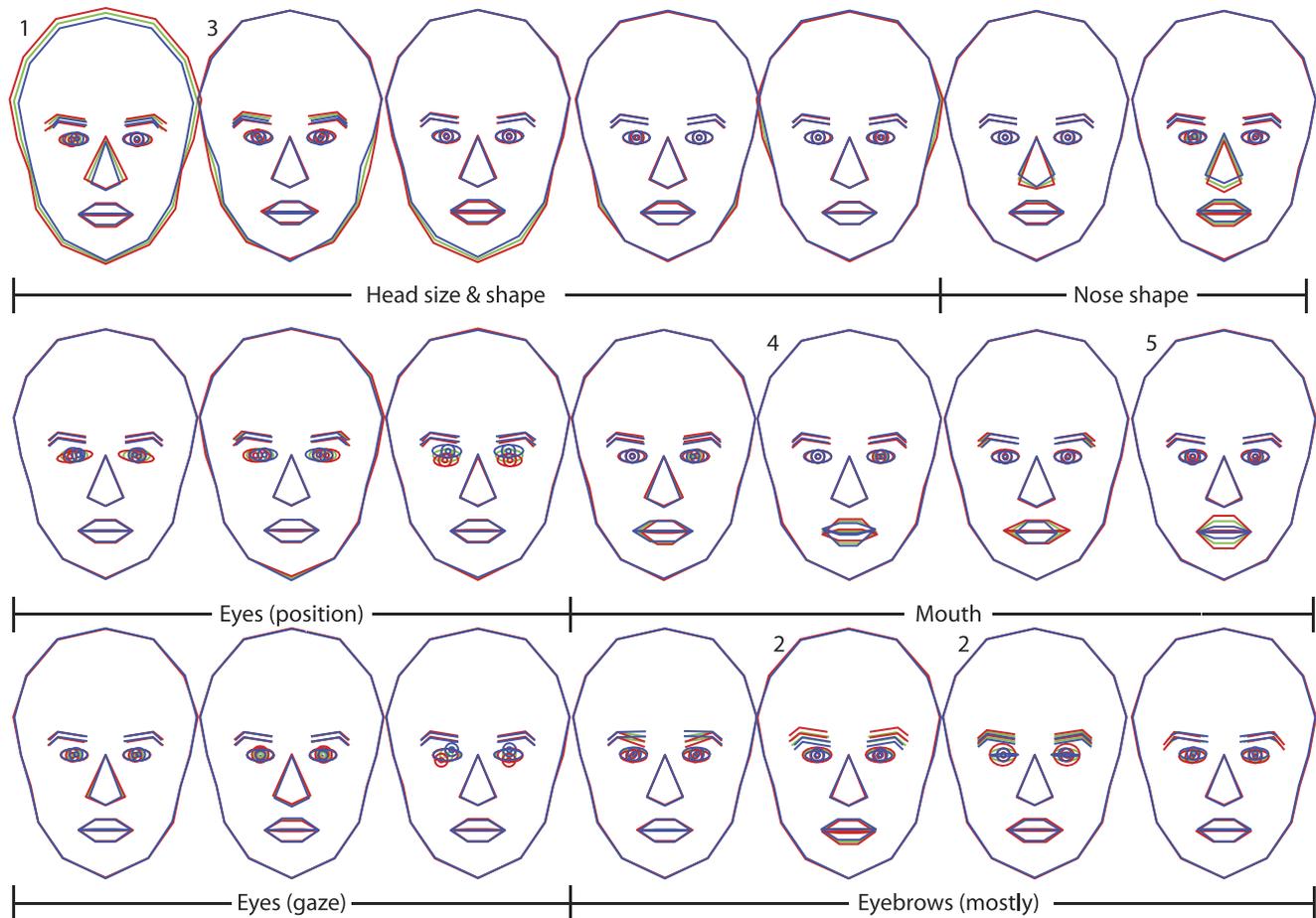


Figure 8. Factor analysis loadings. The loadings for the 21-factors solution. Of these 21 factors, five modify the head outline, three modify the nose, six modify the eyes, four modify the gaze, six modify the mouth, and eight modify the eyebrows (factors including more than one feature type were classified subjectively). The analysis extracted highly symmetric factor loadings, despite most dimensions allowing for asymmetry (e.g., head shape, mouth shape, eyebrow height, eye openness). Even though most factors were local, some included features found at large distances over the face, e.g., head shape, mouth and eyebrow height. Five factors of the six-factor solution were very similar to factors found in the 21-factor solution (labeled 1 through 5 in figure, sixth factor not shown). The two factors shown with number 2 were, in fact, correlated in the results (Figure 9D) and thus integrated as one in the six-factor solution. See text for details.

emotions (pain, disgust, anger, sad). These groupings parallel those found in the previous distance analysis.

Factor analysis

The goal of factor analysis was to determine which facial features covary in groups. Featural information may be integrated in templates to reduce redundancy, compress data, and increase reliability. Factor analysis provides a meaningful grouping of variables based on covariance within the data set, which would be useful for integrative templates.

Factor analysis was performed on raw data. No facial dimension could be predicted from a combination of other facial dimensions (rank = 53 of 53 dimensions). The 21-factor solution is shown after a

varimax rotation (orthogonal); however, a promax rotation (nonorthogonal) produced very similar results. A six-factor solution after a varimax rotation produced five factors very similar to the 21-factor solution (indicated in Figure 8 by numbers) and one factor that included the top of the head, as well as nose and eyebrow shape (not shown). Two factors of the 21-factor solution combined into one in the six-factor solution (see following). Otherwise, the six-factor solution highlights the more important factors in the 21-factor solution. The six-factor and 21-factor solutions provide compression rates of 88.7% and 60.4%, respectively. Figure 9 shows the distribution of results along combinations of some of the factors from the 21-factor solution. The issue of separability of groups has already been addressed with discriminate analysis (see previous), and will only be discussed qualitatively here.

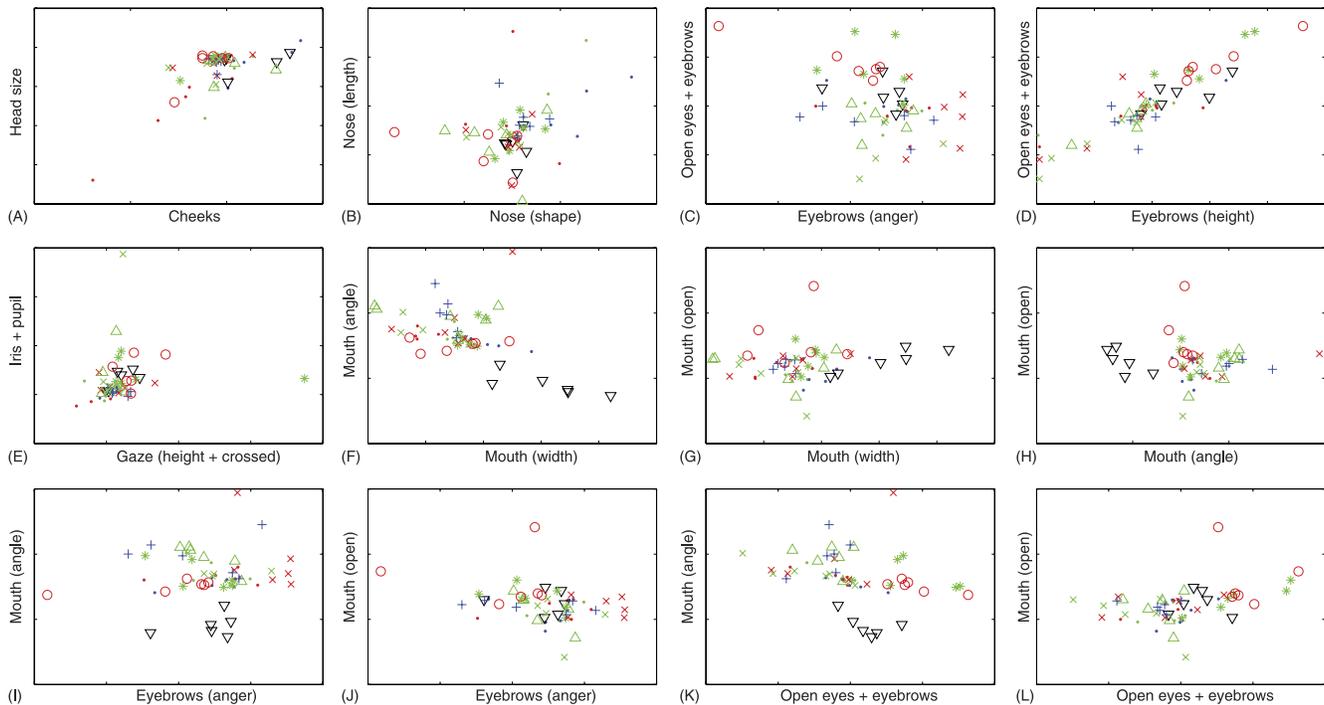


Figure 9. Spread of faces along pairs of factor analysis dimensions (see Figure 8 for dimensions). Symbols are neutral (green dot), male (blue dot), female (red dot), happy (black triangle), afraid (green '*'), angry (red 'x'), disgust (green triangle), sad (blue '+'), surprise (red 'o'), and pain (green 'x'). Although factor analysis does not optimize for separability, some factors do provide some clues toward the dimensions that provide good clustering or separation. The factors for head size and cheeks provide a good separation of genders (A). Mouth angle and width provide a good separation of happiness (F). Note that the factors “iris and pupil size” and “gaze (height + crossed)” shown in (E) are driven by one outlier each. See text for details.

Five factors are related to head size and shape. The first factor is most closely related to head size, but also includes nose width, eyebrow thickness, outer eyebrow length, and eye separation. The second factor is mainly related to cheeks (WSW and ESE mainly) but also includes eyebrow thickness and, to a lower extent, lip thickness and mouth width. Both of those factors are related to gender (see Figure 9A) and also appear in the six-factor solution. The other three factors are related to chin and head shape: chin (SW to SW), jaw (SW and SE), and head sides (W, WNW, E, and ENE).

Nose shape varied in 6 factors: (1) nostril height (nose shape) and lower lip thickness, (2) nose height and length linked to mouth height, eyebrow thickness, and inner length and, to a smaller extent, left-right gaze, (3) nose width linked to head and eyebrows (see previous), (4) nose width with mouth alignment (see following), (5) nose shape with pupil size and mouth openness (see following), and (6) nose width with left-right gaze shifts (see following). Nose characteristics did not seem to provide good separability of faces.

Eye position varied in nine factors: (1) eye angle, width, and gaze crossed/uncrossed, (2) eye and outer eyebrow distance from face center, (3) eye height and width as well as eyebrow length, (4) eye separation with head size, nose width, outer eyebrow length, and

eyebrow thickness (see previous), (5) eye width and openness linked to eyebrow height and separation as well as mouth height (see following), possibly related to scaling of facial inner features, (6) eye openness linked to eyebrow height (see following), (7) eye height linked to outer eyebrow length (see following), (8) eye width and height with crossed/uncrossed gaze (see following), and (9) eye height with inner eyebrow height (see following).

Eye gaze varied in six factors: (1) left-right direction with nose width, possibly related to fear and/or sad, (2) iris and pupil size with lip thickness, (3) height and crossed/uncrossed with eye width, (4) crossed/uncrossed linked to eye angle and width (see previous), (5) right-left gaze with nose length and height, mouth height, eyebrow thickness, and inner length (see previous), and (6) pupil size with nose shape and mouth openness (see following). However, gaze factors 2 and 3 are mainly capturing variance in two outliers in the data, e.g., large deviations that rarely occurred (Figure 9E). Eye position and gaze did not seem to provide good separability of faces.

Mouth characteristics varied in nine factors: (1) mouth horizontal alignment with nose width, (2) mouth angle, e.g., smile, (3) mouth width linked to outer eyebrow length, (4) mouth and teeth openness with

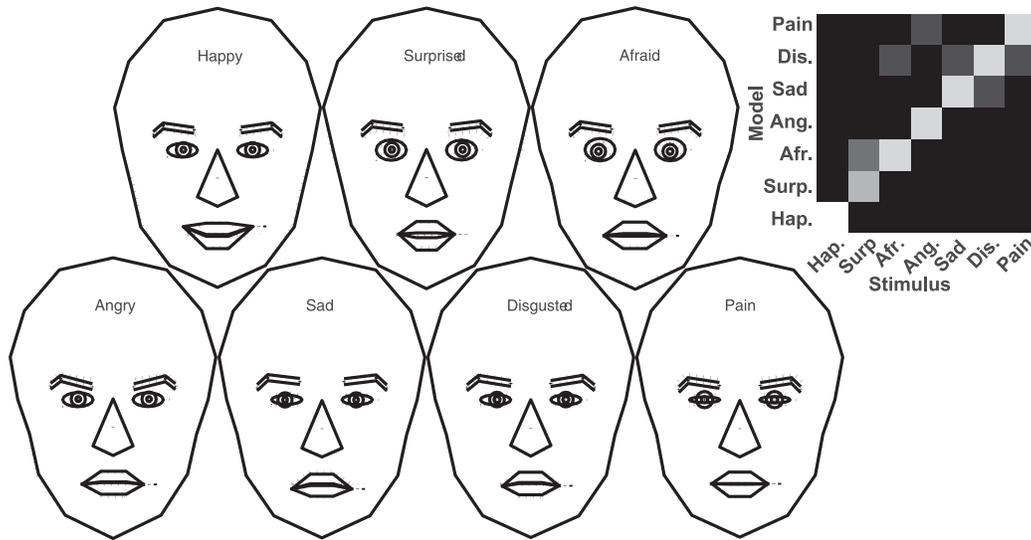


Figure 10. Unique features of each condition, based on dynamic information alone. The classification algorithm was performed data normalized for head size. The classification algorithm achieved 83.3% correct classifications. Much of the emotional expression is carried by the eyebrows, the eyes' openness, and the mouth. Features associated with given emotions include (1) widening and upward movement of mouth corners for happy, (2) teeth opening for surprise, (3) gaze shifting down for fear, (4) downward movement of inner eyebrows for anger, (5) nose shape for disgust, and (6) iris size increase for pain. Sad was not significantly associated with a unique feature. See text for details.

nose shape and pupil size, (5) mouth height covaried with nose length (see previous), (6) mouth height covaried with eyebrow height (see following), (7) lower lip thickness and mouth width with cheeks and eyebrow thickness (see previous), (8) lower lip thickness with nose shape (see previous), and (9) lip thickness with pupil and iris size (see previous). Factors 2 through 4 (mouth angle, width, and openness) provide good clustering of emotions, in particular of happy, surprise, and sad (Figure 9F through H).

Eyebrows varied in 10 factors: (1) inner eyebrow height (e.g., frown) with eye height, (2) eyebrow height and separation linked to mouth height and eye width and openness, possibly related to scaling of facial inner features, (3) eyebrow height and eye openness, possibly related to eye motion in surprise, pain, and fear, (4) outer eyebrow length and eye height with inner eyebrow length, (5 and 6) size and thickness as related to head size and cheeks (see previous), (7 and 8) outer eyebrow length linked to eye horizontal position and to mouth width (see previous), (9) eyebrow thickness and inner length linked with right-left gaze, nose height and length, and mouth height (see previous), and (10) eyebrow length with eye separation and width (see previous). Factors 2 and 3 are correlated (Figure 9D; also see Figure 8 faces labeled with 2). Factors 1 and 3 provide good clustering of surprise, disgust, and anger emotions (Figure 9C).

Combining information across separate features provides better separability. Figure 9I through L show combinations of one mouth-related factor with one

eyebrow-related factor. These combinations provide good clustering of happy, surprise, anger, and fear emotions, particularly when mouth angle is included.

Emotional faces (dynamic)

Based on normalized faces, this analysis uses the difference between ending state and starting state. Static dimensions were removed from analysis. Note that because eye height is static, the eye-eyebrow distance is redundant with the eyebrow height, thus eye-eyebrow distance was not analyzed in this section. This analysis thus revealed dynamic information contained in faces as they deviate from their neutral state to their apex. Each emotional dynamic face was analyzed compared to other emotions, using linear discriminate analysis, to extract unique dynamic features. Classification performance of the linear discriminate analysis was 83.3% whether performed on the normalized data or on the raw data.

Happy face characteristics include an increase in cheeks (WSW, SW, SE, and ESE; $ts(5) > 7.6$, $ps < 0.0007$), increase in outer mouth width ($ts(5) = 12.6$, 19.7 , and 16.6 , $ps < 0.0001$ for left, right, and average, respectively), and upward movement of the mouth corners ($ts(5) = 12.5$ and 13.6 , $ps < 0.0001$).

Surprised face characteristics include cheeks moving inward (W, E, and average; $ts(5) = 30.7$, 16.2 , and 27.4 , $ps < 0.0001$), teeth opening ($t(5) = 27.4$, $p < 0.0001$),

and inner-left mouth moving outward ($t(5) = 8.3$, $p = 0.0004$). Near-significant face characteristics include the average inner mouth moving outward ($t(5) = 5.5$, $p = 0.0028$), eyebrow inner length increasing ($t(5) = 4.5$, $p = 0.0063$), and lips opening ($t(5) = 6.4$, $p = 0.0014$). Not significant by our criterion were eyes opening ($ts(5) = 3.3$ and 3.4 , $ps = 0.022$ and $.019$) and upward movement of eyebrows ($ts(5) = 4.3$ and 4.2 , $ps = 0.0079$ and 0.0089).

Fear face characteristics include chin moving down (SSW, S, and SSE; $ts(5) = 15.9$, 6.0 , and 12.6 , $ps < 0.0001$, $= 0.0018$ and < 0.0001) and gaze shifting down ($t(5) = 18.5$, $p < 0.0001$). Near-significant results include eyes opening ($ts(5) = 5.3$ and 4.9 , $ps = 0.0033$ and 0.0046).

Anger face characteristics include inner eyebrows lowering ($ts(5) = 16.5$ and 14.8 , $ps < 0.0001$) and inner-right mouth widening ($t(5) = 12.2$, $p < 0.0001$). Near-significance characteristics include upward movement of outer eyebrows ($ts(5) = 6.6$ and 6.5 , $ps < 0.0013$), average inner mouth widening ($t(5) = 5.8$, $p = 0.0021$), average mouth height increasing ($t(5) = 6.2$, $p = 0.0016$), and the SW jaw tightening ($t(5) = 5.0$, $p = 0.0043$; SE jaw not significant by our criteria: $t(5) = 4.5$, $p = 0.0065$).

Sad face characteristics did not include statistically significant components at the criterion used here. Near-significant face characteristics include only gaze shifting left ($t(5) = 6.1$, $p = 0.0017$). Not significant by our criteria include downward motion of eyebrows ($ts(5) = 4.2$ and 4.6 , $ps = 0.0085$ and 0.0057) and mouth corners ($ts(5) = 2.82$ and 2.78 , $ps = 0.037$ and 0.039).

Disgust face characteristics include only the nose shape ($t(5) = 7.1$, $p < 0.0009$). Near-significant face characteristics include outer eyebrow length increase ($t(5) = 5.7$, $p = 0.0034$) and gaze alignment crossing ($t(5) = 5.1$, $p = 0.0037$).

Pain face characteristics include only iris size increase ($t(5) = 23.1$, $p < 0.0001$). Near-significant face characteristics include left eyebrow height decrease ($t(5) = 5.1$, $p = 0.0037$; right was not significant: $t(5) = 4.6$, $p = 0.0056$), eyebrow outer length decrease ($t(5) = 6.5$, $p = 0.0013$), and pupil size increase ($t(5) = 5.1$, $p = 0.0039$).

Neutral

The neutral face is often assumed to act as a reference or origin in face space. However, there exists multiple ways to define a gender-neutral or emotion-neutral face. By including a neutral condition, we have a template of what participants believe a “neutral face” looks like, named “measured neutral” in this section to distinguish it from other definitions of the neutral. Here, we compare this measured neutral with morph faces composed of weighted sums to assess whether

different definitions of “neutral” are equivalent or dissimilar. The Euclidian distance between faces was used as a measure of similarity. In particular, we compare this measured neutral to two other definitions of neutral: (1) the *average morph*, i.e., the central tendency of the distribution of related conditions and (2) the *optimal morph*, i.e., the weighted sum of conditions to create the closest morph to the face given in the neutral condition.

Gender morphs were created by linear interpolation between the average male and average female faces (normalized for head size). The optimal morph was 38% male, 62% female face. This deviation from 50%–50% is not consistent across participants ($t(5) = .32$, $p = 0.77$) with some participants’ measured neutral being closer to their male or female condition. Male and female faces are on average about equally distant to the measured neutral face ($t(5) = 1.1$, $p = 0.31$). Similarity to the measured neutral differed between the four points sampled on the male-female continuum ($F(3, 15) = 5.5$, $p = 0.01$) with the measured neutral closer to the average morph and to the optimal morph than to either pure gender. The optimal morph was consistently closer to the true neutral than the average morph ($t(5) = 3.15$, $p = 0.026$). Nevertheless, the two morphs (average and optimal) were very similar to the measured neutral. Our data thus support that a 50% morph is a decent approximation of a gender-neutral face for average participants, although it is unclear whether variability between participants reflect measurement noise or true idiosyncrasies.

The optimal emotion-neutral morph was created using the following steps: (1) seven positive weights were generated, (2) these weights were normalized such that their sum equals one, (3) the emotional-morph was defined as the weighted sum of emotional faces, using these seven weights on the seven emotional faces, and (4) gradient descent was used to adjust weights such that the morph created was as similar as possible to the measured neutral, while always keeping the sum of weights equal to 1 and keeping every weight positive. For the static morphs, the statistical calculations were performed using two neutrals: (1) average neutral condition across participants or (2) neutral condition for each participant separately. For dynamic morphs, the only neutral condition corresponds to “no motion.”

Statistical significance of weights was calculated for both static neutrals as well as for the dynamic neutral. The emotions that were included reliably in neutral morphs in at least two of the three calculations were happiness ($ts(5) = 9.67$, 3.68 , and 2.52 , $ps = 0.0002$, 0.014 , and 0.053 for static average neutral, static individual neutral, and dynamic neutral, respectively), sadness ($ts(5) = 3.37$, 3.04 , and 2.61 , $ps = 0.020$, 0.029 , and 0.048), and anger ($ts(5) = 6.64$, 2.71 , and 2.34 , $ps = 0.0012$, 0.042 , and 0.066). Disgust was significant only

in the static average neutral condition ($ts(5) = 3.48, 2.03, \text{ and } 2.32, ps = 0.018, 0.098, \text{ and } 0.068$). All other emotions were not statistically significant for any of the three calculations (all $ts(5) < 2.16, ps < 0.083$), suggesting that the inclusion of fear, surprise, or pain into a neutral morph tends to degrade the quality of that neutral morph. This supports that the measured neutral face is not well-approximated by the average morph face.

Morph faces created using optimal weighted sums of emotions do provide fair approximations of measured neutral faces (especially in the dynamic case). However, some differences remain including: (1) open vs. closed mouth for static and dynamic faces and (2) nose length and nostril height for static faces.

Discussion

Summary of results

Participants made female faces with smaller head sizes, thinner noses, and thinner eyebrows. Once normalized, female faces were more oval-shaped than men's and still had thinner eyebrows and thinner noses.

Different emotions were driven by different parts: (1) happiness by higher and wider mouth corners, (2) surprise by higher eyebrows as well as opening of eyes, lips, and teeth, (3) fear by the opening of eyes and lips, (4) anger by the lowering of inner eyebrows, (5) sadness by lowering mouth corners, sloping eyes, and gaze shift toward the left, (6) disgust mainly by higher inner eyebrows and higher nostrils, but also by mouth corners lowering and eyes closing, and (7) pain by open lips, closing of eyes, size increase of iris and pupil, and lowering of eyebrows. Based on these features, linear discriminate analysis achieved classification accuracy of 75.0% to 83.3%. Some features were infrequently reported, e.g., gaze shift in fear, pupil and iris size changes in pain, eye angle in sadness. Further studies could test whether these infrequently reported features nevertheless contribute significantly to the quality of the percept.

Static and dynamic cues were very similar, which is not surprising given that participants were told to make the expression start at neutral and end at the given expression. Main differences between these analyses were: (1) static-only dimensions were not included in dynamic analysis and (2) the dynamic classification analysis was generally less powerful than its static counterpart.

Classification features were more selective, as shared features are more difficult to use to classify facial expressions accurately. Thus, classification features tended to emphasize features unique to each face

condition. No single feature was significantly associated with sad, including the lowering of mouth corners. Surprise and fear shared the opening of the eyes and lips, so those cues were not significant for the purpose of classification. Disgust could not be classified reliably based on the lowering of mouth corners (shared with sadness) or eyes closing (shared with pain). On the other hand, classification did reveal some features that were either too subtle or too variable, such that these features were only detectable given a larger context of faces across a larger number of categories: upward nostrils for disgust, sloping eyes for sad, lower eyebrows for pain (static), pupil and iris size increase for pain.

Similarity was measured using distance (Euclidian sum of geometric distances) and correlations (similarity of geometric changes independent of intensity). Happiness is unique, as it holds a position in space that is far away from other emotions. Surprise is also far away from other emotions except for fear, which has a similar expression to surprise but at a lower intensity. Neutral is the face closest to other faces on average. Pain, anger, disgust, and sad were closer to each other than to other faces.

Factor analysis revealed that correlations between feature dimensions were usually fairly local (i.e., contained within the same feature) except for a high level of symmetry. Nevertheless, there was some dependency between spatially-distant features, including: (1) head size with internal feature sizes, similar to gender-related changes, (2) internal feature position scale, including eyes, eyebrows, and mouth, and (3) eyebrow height and eye openness, similar to eye motion in surprise, pain, and fear (see Factor analysis section of [Results](#) for more details).

A 50% morph between female and male faces is a decent approximation of a gender-neutral face. The emotion-neutral morph that best approximates the true neutral face is composed of happy, anger, sad, and maybe anger as well. An average of all emotional faces produces a poor neutral. Therefore, to produce a neutral emotional face using a weighted sum approach, some emotions are better left out.

Linear morphing techniques are often used to generate a dynamic transition between emotional expressions obtained from photographs. Dynamic trajectories did not significantly deviate from linear trajectories on average, meaning that a linear morph made from the starting to the ending state would be an acceptable approximation of intermediate states. Of the 42 emotional faces collected from participants, 38.1% of dynamic trajectories were linear. While it is still unclear whether nonlinear trajectories could further improve the expression of emotions, it is clear that participants did not feel the need to make their emotional faces nonlinear in order to make them look optimal.

Features: gender

Many studies focus on describing the features involved in given discrimination tasks, e.g., gender or expression discrimination. Our analyses of features directly compares with literature on parts or features.

Gender discrimination is mostly supported by information around the eyes, the eyebrows, the jaw, the nose, the mouth, and the face outline (e.g., Brown & Perrett, 1993; Burton, Bruce, & Dench, 1993; Campbell, Benson, Wallace, Doesbergh, & Coleman, 1999; Dupuis-Roy, Fortin, Fiset, & Gosselin, 2009; Gosselin & Schyns, 2001; Mangini & Biederman, 2004; Russell, 2003, 2005, 2009; Schyns, Bonnar, & Gosselin, 2002; Yamaguchi, Hirukawa, & Kanazawa, 1995). Brown and Perrett (1993) isolated individual features, which they then asked participants to classify in isolation or grafted onto the opposite gender to measure the importance to gender of each cue. Their features carrying gender information were (in descending order): brows and eyes, brows alone, eyes alone, whole jaw, chin, nose and mouth, and mouth alone. The nose did not carry gender information. The features affecting gender within the context of a face were (in descending order): jaw, brows and eyes, chin, and brows.

Our results show that head size is an important cue to gender. Whether normalized for head size or not, female faces were more oval-shaped than men's and had thinner eyebrows and thinner noses. Gender was recognizable from average male and female faces and showed similarities to Wilson's measured gender differences. Our method also produced an additional cue, head size, which is usually missed in other studies where faces are normalized for head size.

In our results, the eye-eyebrow distance was very similar across gender. This conflicts with claims that the eye-eyebrow is most reliable relational cue (Burton et al., 1993; Campbell et al., 1999). However, neither Burton's or Campbell's study is conclusive. Burton et al. (1993) did not measure human performance; rather, they measured facial properties of faces and used discriminant function analysis. Their recovered features are thus ideal for computer-based gender discrimination, but it is not clear that humans use these ideal features.

Campbell et al. (1999) manipulated eye-eyebrow distance and measured gender classification bias. They manipulated eye-eyebrow distance by having actors (hired to produce stimuli for the study) frown or raise eyebrows, look up or down, and angle their head up or down. These manipulations changed eye-eyebrow distances; however, they also included systematic changes in many extraneous dimensions including eyebrow angle, head angle, and gaze direction. It is thus not clear that participants' bias was due to eye-

eyebrow distance. Even if we accept the author's claim, the eye-eyebrow effect was small (significant in three of four tests, with its most significant at $p = 0.02$). Moreover, the eye-eyebrow distance effect was not replicated in a study using bubbles (Taschereau-Dumouchel, Rossion, Schyns, & Gosselin, 2010). A hypothesis that could reconcile the different data would be to say that, while the cue may be present in stimuli and potentially useful, humans do not use this cue when presented with other cues to gender. However, further research is needed to clarify the importance of this gender cue.

Features: emotions

Similar to gender, facial expressions are also concentrated in few features. Gota and Miyamoto (2000) showed that the top of the face was more important for anger, fear, surprise, and sadness, whereas the bottom was more important for disgust and happiness (see also Bassili, 1978, 1979; Boucher & Ekman, 1975; Butler, Blais, Gosselin, Bub, & Fiset, 2010; Caldara et al., 2005; Ellis, Shepherd, & Davies, 1979; Schyns et al., 2002; Young, Hay, McWeeny, Flude, & Ellis, 1985). At a finer scale, Smith, Cottrell, Gosselin, and Schyns (2005) measured the areas of facial expressions where useful information was present for categorization using the bubbles technique. Human responses showed that anger, fear, and sadness used different features than other emotions, and that the high similarity of fear and surprise was lowered by focusing on different features. The mouth was useful for happy, surprised, and disgusted, whereas the eyes were useful for fear and anger. Sad was distributed. Using similar masking methods, other researchers have confirmed the importance of the mouth for happy (Gosselin & Schyns, 2001; see also Schyns et al., 2002). Using classification image methods, Mangini and Biederman (2004) found eyebrows and mouth for happy vs. unhappy. Actors also show distributed facial movements when acting pain (Simon, Craig, Gosselin, Belin, & Rainville, 2008).

In our data, emotion information was also distributed differently for different emotions: (1) happiness in the mouth region, (2) surprise was distributed across the face, e.g., eyebrows, eyes, and mouth, (3) fear was mostly in the eyes, but also in the mouth, (4) anger was in the eyebrows and mouth, (5) sad was distributed in the mouth and eye regions, (6) disgust was distributed, e.g., eyebrows, eyes, mouth, and nose, and (7) pain was distributed, e.g., eyebrows, mouth, and eyes. Overall, our results replicate previous results, except that our data also identifies additional features thus giving a more distributed pattern of features. This is likely a natural consequence of our method being sensitive at

suprathreshold levels, thus enabling more subtle features to emerge compared to more difficult tasks where only the most salient task-relevant features tend to be significant in the results. That is, our method finds features that are present in the templates, including features that may not be necessary to task performance.

Features: integration

Integration of features has been studied mainly under “holistic processing,” defined as a stronger and automatic perceptual integration across facial features, even distant ones, especially for precise spatial-relational information and feature shape (Cheung, Richler, Palmeri, & Gauthier, 2008; Farah, Wilson, Drain, & Tanaka, 1998; McKone et al., 2007; Robbins & McKone, 2007; Tanaka & Farah, 1993; Tyler & Chen, 2006; Young, Hellawell, & Hay, 1987; Yovel & Kanwisher, 2005; see review in Maurer, Le Grand, & Mondlock, 2002).

Much of the evidence for holistic face processing comes from the face inversion effect where inverted faces are processed less efficiently, possibly due to a disruption of holistic processing (Farah, Wilson, Drain, & Tanaka, 1995; Leder & Carbon, 2004; Moscovich & Moscovich, 2000; Russel, Duchaine, & Nakayama, 2009; Van Belle, De Graef, Verfaillie, Rossion, & Lefèvre, 2010; Wilson et al., 2002; Yin, 1969; see reviews in Rossion 2008, 2009; Valentine, 1988, 1991). Further evidence includes performance impairments due to task-irrelevant changes to the stimuli, e.g.: displacing features (Tanaka & Sengco, 1997), changes in face information (Cheung et al., 2008; Farah et al., 1998; Gauthier & Tarr, 2002; Richler, Bukach, & Gauthier, 2009; Richler, Gauthier, Wenger, & Palmeri, 2008; Tanaka & Farah, 1993; Wenger & Ingvalson, 2002) and misaligning top and bottom of faces (Cheung et al., 2008; Gauthier & Bukach, 2007; Goffaux & Rossion, 2006; Richler et al., 2008; Richler, Tanaka, Brown, & Gauthier, 2008). There is also some evidence that encoding of feature position is dependent on distant features (Barton, Keenan, & Bass, 2001; Goffaux & Rossion, 2007; Malcom, Leung, & Barton, 2005; Sekunova & Barton, 2008).

The holistic processing account is still under considerable debate (e.g., Konar, Bennett, & Sekuler, 2010; Sekuler, Gaspar, Gold, & Bennett, 2004; Taschereau-Dumouchel et al., 2010). In particular, even if we accept the holistic account, we still do not know which features are integrated together. Our study provides a number of specific feature combinations that are likely candidates for integration.

Factor analysis in our dataset has shown several factors where information was correlated across facial features located in different regions, including (1) head

size with internal feature scaling mostly along the horizontal axis, (2) cheeks with thickness of lower lips and eyebrows, and (3) internal feature scaling mostly along the vertical axis as well as feature thickness (see Factor analysis section of Results). These correlations were present in the templates, and we suspect they would be present in natural statistics of faces as well. Gender is associated with a set of correlated features and so are many emotions. It is thus likely that long-range face mechanisms would integrate over these specific combined feature changes. Also, our method can be used to generate stimuli where these correlations are manipulated to assess holistic processing. These targeted manipulations may prove more sensitive than the more traditional techniques used so far.

Face space

Theoretical and empirical methods on the internal mapping of face representations refers to this representation as “face space.” The more common face space theory states that the average face is the origin of the face space (e.g., Blanz, O’Toole, Vetter, & Wild, 2000; Leopold, O’Toole, Vetter, & Blanz, 2001; Loffler, Wilkinson, Yourganov, & Wilson, 2004; Valentine 1991; Wilson et al., 2002) and emotions are organized in roughly polar opposites (e.g., Plutchik, 1980, 2001). However, there is place for debate. For example, Rutherford, Cattha, and Krysko (2008) argued that positive emotions are lumped into one category (happy) whereas negative emotions are distinct. Here, we review the literature regarding face space and discuss the relevance of the current study.

Face space: layout

We used distance and correlations as our measures of similarity between facial expressions. By integrating local differences into global measures, we sought to better understand the face space. We determined that happiness was unique, as it was further from other emotions. Surprise was also far away from other emotions except for fear. Fear was similar to surprise but at a lower intensity. Other emotions differed from each other in more subtle ways.

Goren and Wilson (2006) measured threshold discrimination performance as well as classification performance for four emotions: happy, sad, fear, and anger. Their discrimination task consisted of reporting whether a viewed face was neutral or had an effect, and thresholds were measured as deviations in physical units rather than proportion of morphing. They found that participants were more sensitive to fear (i.e., best threshold), followed by happy, then sad, and finally

sensitivity was worse to anger. Their 5AFC classification task (four emotions + neutral) showed that emotions were ranked as follows: happy (easiest), fear and sad about equal, and anger (hardest). Fear and sadness were often confused.

Our study measured the total distance between faces, and theirs measured the shortest distance from neutral toward a given face that can be discriminated from neutral. To compare results across these two studies, we assumed that shorter distances between two face categories would make the faces more difficult to discriminate, an assumption commonly found in face perception models. This assumption predicts that the easiest to hardest faces to discriminate from neutral, based on our results, would be surprise, happiness, fear, pain, sad, disgust, and finally anger; thus the four emotions used by Goren and Wilson should rank as happiness, fear, sad, and finally anger. That is, our results are consistent with Goren and Wilson's (2006) 5AFC results and similar to their discrimination task except that fear ranks lower than happiness in our results. On average, easily classifiable emotions in their study were further from other emotions in ours.

The similarity of results across very different paradigms suggest that they both measure a similar layout, where happy is unique, surprise and fear are similar but of different intensity, and negative emotions are more similar to each other than to other emotions.

Face space: origin

Many theories of face space assume that the average face is the true origin of face space and that faces are represented as deviations from that origin (e.g., Loffler, Yourganov, Wilkinson, & Wilson, 2005; Rhodes, 1996; Rhodes et al., 1987; Valentine, 1991; Valentine & Bruce, 1986). There is some supporting evidence from the face adaptation literature for this assertion: (1) adapting to an average face does not shift the percept of other faces (Webster & MacLin, 1999), (2) adapting to an average face does not affect the discrimination curve for identity (Leopold et al., 2001), and (3) face adaptation effects are generally smaller if adapt and test stimuli are sampled on a line that does not pass through the average face (Anderson & Wilson, 2005; Leopold et al., 2001; Rhodes & Jeffery, 2006). Other evidence for the specialness of the origin includes: (1) discrimination is harder if the two stimuli to be discriminated are located further from the average face (Dakin & Omigie, 2009; Wilson et al., 2002; but see Rhodes, Maloney, Turner, & Ewing, 2007) or are sampled along a line that does not pass through the average face (Wilson et al., 2002), (2) face identity cancels more if a face is combined with its antiface than if it is combined with any other face (Rhodes & Jeffery,

2006), and (3) distinctive faces are also further from the average face (Johnston, Milne, Williams, & Hosie, 1997). These studies looked at identity or geometric distortions of faces, not at emotions.

Adaptation has been found for facial expressions and is generally consistent with positive and negative emotions being opposites of the neutral or average face. Supporting this, adapting to a positive (or negative) facial expression makes the neutral appear more negative (or positive; e.g., Rutherford et al., 2008). Also, adapting to the opposite of an emotion (i.e., antiface) makes the neutral appear more like the original face (Skinner & Benton, 2010; Webster et al., 2004), e.g., adapting to the antiface of sad makes a neutral face appear sad. However, unlike with identity or distortion, adaptation effects for facial expressions don't need to be sampled along a line that passes through neutral. Adapting to an emotion biases the percept of ambiguous stimuli away from the adapting stimulus seemingly along any axis sampled, including axes defined by combinations of happy, sad, angry, disgusted, surprised, and fear (Benton et al., 2007; Ellamil, Susskind, & Anderson, 2008; Fox & Barton, 2007; Hsu & Young, 2004; Webster et al., 2004). Many axes for which a significant adaptation effect was found did not pass through either the neutral or the average face. Therefore, there does not seem to be anything special about the origin of face space for facial expression perception.

Whether the neutral or average face is the origin of face space, the commonly-held assumption has been that the average face and the neutral face were the same. The terms are used interchangeably in the literature (e.g., gender-neutral, gender-average, neutral emotion, average face). The only time these terms differ is to denote methodological or quantitative differences, e.g., the "neutral" is usually a resting state (for facial expression) or a point of subjective equality (for gender) rather than a mathematical average. Even then, no evidence has been presented to suggest that there is a fundamental difference between the average and the neutral.

In our study, however, the average face is quite different from the neutral face over our sampling of facial expressions. Indeed, within the first two PCA dimensions, the neutral face is away from the center of the facial expression space, much closer to the edge of that space (Figure 7, bottom). The main feature behind that discrepancy is mouth openness. The mouth is closed in the neutral condition, from which position it can only open. By contrast, the mouth is opened in the average across emotional conditions, because several emotions include an open mouth (or at least open lips) including happy, surprise, and fear. Any average of faces that include a fair sampling of facial expressions

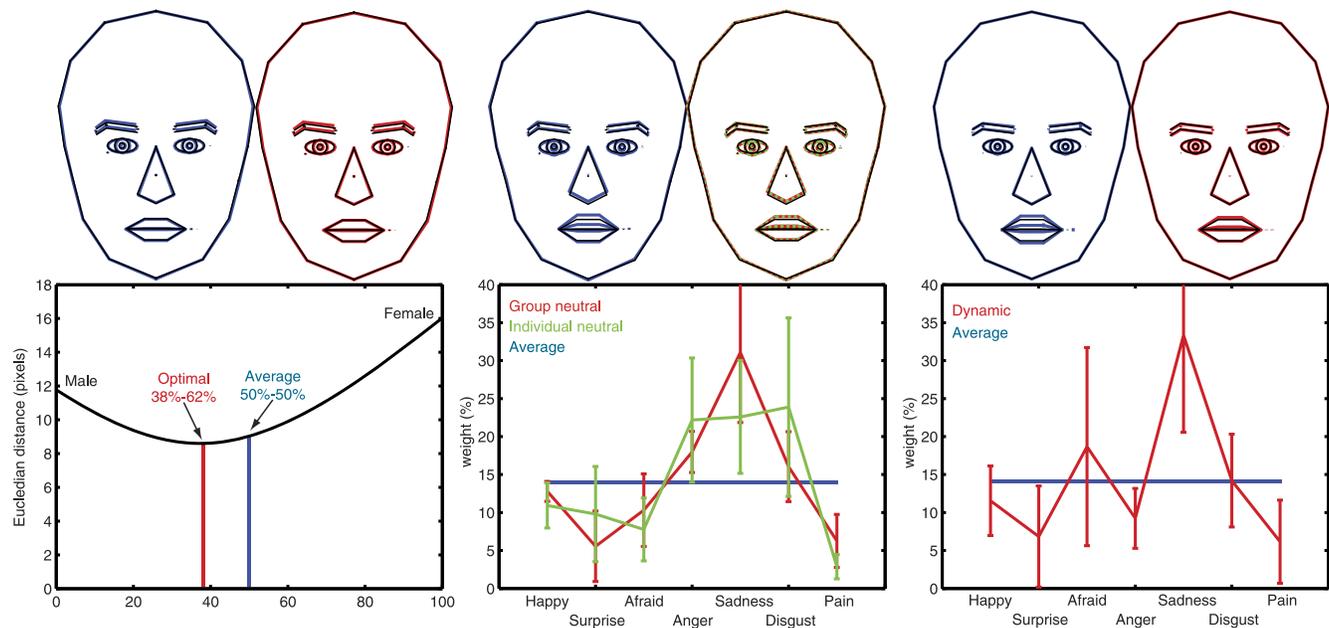


Figure 11. The neutral face derived experimentally (“measured neutral;” black) compared to neutral faces created using morphs of other conditions (average in blue, optimal in red or green), for gender (left), static emotion (intermediate), and dynamic emotion (right). Shown are the various neutral faces (top) and the weights used to construct them (bottom). (Left) Measured neutral vs. faces morphed to be gender-neutral faces. The face on the male-female line closest to the measured neutral face is 38% female, 62% male. (Intermediate) Measured neutral vs. faces morphed to be emotion-neutral for static faces. Two optimal morphs are shown: morphs obtained as the closest fit to individual neutrals (green) or to the group neutral (red). (Right) Measured neutral vs. faces morphed to be emotion-neutral for dynamic faces. For both static and dynamic faces, the face in emotional space that is closest to the neutral face is a combination of happy, anger, and sadness, and to a lesser extent, disgust.

(even excluding verbal communications) will likewise include an open mouth.

Face space: importance of origin choice

The choice of average or neutral as origin will impact the construction of antifaces (and whether such antifaces are physically plausible) as well as model predictions near the origin of face space. We argue here that earlier studies used methods that were not sensitive to this choice of origin for the following reasons. First, it is possible that most features are roughly normally distributed with the neutral and average being similar. Mouth openness is likely a rare exception. Second, even if a dimension does not have negative values, effects such as adaptation can nevertheless occur. For example, adaptation to an open mouth could make it harder to detect a small mouth opening. Supporting this, adaptation effects occur to mouth openness in both directions (i.e., from opened to closed, and from closed to opened; Jones, Feinberg, Bestelmeyer, DeBruine, & Little, 2010). Third, as discussed previously, adaptation to facial expression occurs even with stimulus sampling lines that do not go through the average or neutral face. Fourth, adaptation to an

expression may have a similar effect on the average and the neutral face. This was emphasized by Cook, Matei, and Johnston (2011), stating “one can conclude only that perception is being shifted away from the adapting stimulus; the precise direction of the shift remains ambiguous” (p.2). In summary, study results on the facial expression face space may be robust with respect to choice of origin, in large part because these studies were not designed to be sensitive to that choice.

Figure 12 shows antifaces constructed with the neutral condition used as origin for the face space, for each of the nine non-neutral conditions, using the method described by Rutherford et al. (2008) “to artificially construct the perceptual ‘opposite’ of a facial expression, by moving each feature back to neutral and then beyond by as many pixels” (p.40). The resulting antifaces are fairly easy to recognize as human, opposite to Rutherford’s claim that “the resulting face would look nothing like the emotional opposites found in these experiments. Indeed, it would not look like a human facial expression of any sort” (p.40). Moreover, the antiface of male looks female, and the antiface of female looks male. This is consistent with research showing that genders are on opposite sides of the neutral (e.g., Webster et al., 2004). Also, aside from the impossibility of mouths being overclosed (shown in Figure 12 as lips overlapping), the antifaces of

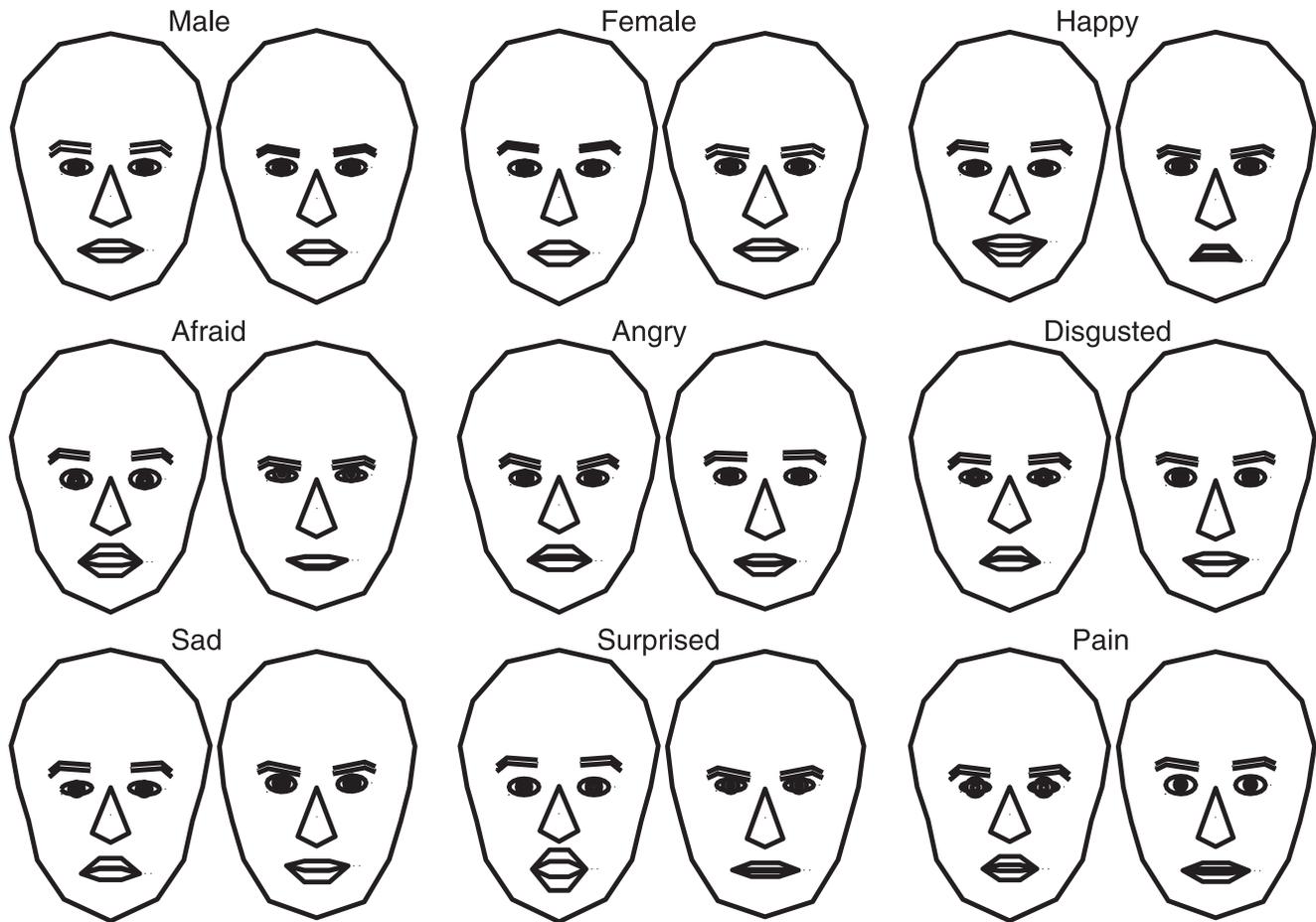


Figure 12. Average faces for each non-neutral condition (left of each pair) and their antiface version (right of each pair). The antifaces share many similarities with their opposites, e.g., the antiface of male looks female. While the antiface of negative emotions tend to look happy, the specific type of happiness differs (e.g., the antiface of angry is peaceful and the antiface of afraid is defiant). Likewise, the antiface of happy is unhappy, which contains elements of several negative emotions (e.g., sadness, anger). See text for details.

emotional faces look like recognizable emotions. The antifaces of negative emotions can be broadly classified as positive (happy), and the antiface of our positive emotion (happy) appears negative (see following for more details). Thus, positive and negative emotions appear to be on opposite sides of the neutral. Similar results are obtained if using the average face instead of the neutral face as the point of origin, supporting the robustness of results across choices of origin.

Face space: subdivisions of happy

Rutherford et al. (2008) suggest that positive emotions are all conveyed through the facial expression “happy.” They support this with experiments where adapting to a negative emotion (i.e., anger, fear, surprise, disgust) makes a neutral face appear happy. Although one experiment allowed free responses, they acknowledge the “subjective nature of the post hoc

categorization of the responses” (p. 31). They included a table showing how individual responses were categorized. Inspecting that table, we note that, of 48 responses they classified as happy, 15 responses were “relaxed,” six were “content,” and four were “satisfied.” Moreover, their “objective” experiment allowed only one positive emotion (happy) contrasted to four negative emotions (sad, fear, anger, disgust). Although they conclude that adapting to any of the four negative emotions produces a happy aftereffect, we feel that their results are best summarized in their own words: “each of the four negative emotions may be producing a unique aftereffect, but given the six alternatives, happiness is the best label for each of these unique aftereffects” (pp. 39–40). In other words, their method lacked sensitivity to subcategories of positive emotions.

Some results from our study suggests that “happy” can be subdivided into different facial emotions, consistent with Plutchik’s (1980, 2001) theory that negative emotions have matched positive emotions (i.e.,

joy-sorrow, anger-fear, acceptance-disgust, surprise-expectancy). Looking at data for individual participants, one can readily see several kinds of happiness (Figure 4). That is, different participants created happy faces that differed in intensity and quality, some ecstatic, some content, and some happy. Further supporting this, the antifaçes of negative emotions are more specific than just “happy” (Figure 12); the antifaçe of angry looks peaceful or content, the antifaçe of fear looks courageous or defiant, and the antifaçe of disgusted looks interested. Only the antifaçe of sad is hard to describe as something else than happy. Also, the antifaçe of surprised can be interpreted as sleepy or annoyed. Although the current study was not designed to look for different kinds of happiness, it appears to have the proper tools to detect these subtle differences. A follow-up study could provide more specific labels (e.g., “ecstatic,” “laughing,” “content,” “peaceful”) which could lead participants into producing faces more specific to each category. Thus, our method could be used to avoid possible categorization biases present in other studies (e.g., Rutherford et al., 2008).

Other synthetic face methods

Other researchers have used synthetic faces to study facial expressions (e.g., Brunswik & Reiter, 1937; Garrod et al., 2010; Goren & Wilson, 2006; Kostov et al., 2001; Sigala & Logothetis, 2002; Wilson et al., 2002). Synthetic faces are generally easy to morph, providing simple controlled ways to generate new faces. As all facial dimensions in our stimuli are expressed in spatial units, our stimuli are ideally suited to morphing.

The stimuli used by Wilson et al. (2002; see also Goren & Wilson, 2006; Loffler Gordon, et al., 2005) offers 37 dimensions, compared to the 53 dimensions used here. They opted for stimuli that offered better controls over spatial-frequency content, using bandpass edges instead of more naturalistic line or edge information, even though natural edge information is important (Bruce, Hanna, Dench, Healey, & Burton, 1992). Their stimuli and ours are similar enough in construction, dimensionality, and general appearance that results should generalize across stimuli. The main difference between Wilson and colleagues’ work and the current study was methodological; they measured threshold performance based on stimuli that matched natural statistics, whereas we had participants adjust facial features subjectively. It is worth noting that, despite methodological differences, we found similarities between our results and Wilson’s, e.g., (1) our neutral condition was similar to Wilson et al.’s (2002) average face, showing that our participants did produce a good representation of a neutral face, (2) their male and female averages were also very similar to our male

and female faces in shape, feature placement, and feature sizes (normalized features correlated well across studies: $r = 0.390$, $t(51) = 3.03$, $p = 0.0039$), and (3) our participants replicated Wilson et al.’s reported average head radius difference between males and females (Wilson et al., 2002: 7%; present study: 4.5%; unnormalized features correlated highly across studies: $r = 0.813$, $t(51) = 9.97$, $p < 0.0001$). That is, despite that our stimuli did not include a priori gender features, our participants produced similar gender characteristics as found in Wilson et al.’s dataset.

Kostov et al.’s (2001) method is most similar to our experimental methodology, although with very different stimuli. They asked participants to create facial expressions using cartoon faces composed of lines and ellipses. Their cartoon faces varied on nine dimensions, far fewer than our 53 dimensions. They also measured the same emotions in real faces. Interestingly, they found that five emotions (sad, tired, scared, troubled, and stress) clustered together in both real and cartoon faces, unlike happy, surprised, anger, and normal. For normal faces, anger was closer to the negative emotions than to happy or surprise. For cartoon faces, however, happy and surprise were closer to other negative emotions than anger was. We found similar clustering of negative emotions (e.g., sad, disgust, pain, and to a smaller extent anger). In our study, anger was closer to negative emotions than happy or surprise, consistent with Kostov et al.’s data on real faces. This further supports that artificial faces are relevant to face perception research provided they are sufficiently detailed. Similarly, sufficiently-detailed cartoon faces affects perception of real faces, suggesting that common mechanisms respond to both types of stimuli (Chen, Russell, Nakayama, & Livingstone, 2010).

Garrod et al. (2010) showed three-dimensional faces with randomly changing parameters to participants, and participants had to rate these stimuli on a number of emotions. Participant’s subjective ratings were then reverse-correlated to extract features associated with emotions, e.g., they estimated which features were correlated with a high “happy” rating. Their method recovered features that were associated with a category, similarly to our method. It is not clear whether their method extracts only unique features (as did our discriminant analysis) or whether it also extracts ambiguous features associated with several categories (as did our regular analysis). One advantage of our method is that both ambiguous and unique features are contained in participant responses; the disambiguation is done at the level of the analysis. The faces produced by Garrod et al.’s method also included considerable amount of noise, which would be expected to disappear with averaging over a large enough number of trials. By comparison, our method uses nonrandom stimuli permitting the recovery of face templates of high

complexity within a reasonable number of trials. The variability in our analyses was mainly due to intersubject differences in templates as our stimuli were noise-free.

Conclusion

We presented here a novel method for measuring internal templates for facial recognition. This novel method was able to recover from participants most of the known features associated with gender and facial expression of emotions, as well as several subtle features. Factor analysis also revealed which features were covaried within the data set, hinting at likely candidates for integration of information across features. Further analyses showed that morphing is a fair approximation in a variety of cases, e.g., (1) creating a neutral-gender from male and female faces and (2) creating a dynamic face from starting and ending states. Our data analysis also forms the first mapping of the internal representation of the face space; other such mappings are either based on natural statistics rather than internal representations or based on theoretical grounds rather than empirical. This mapping shows that happiness is unique, and surprise is unique except for fear being similar to surprise but at a lower intensity.

The method and stimuli generated are useful for general face perception research. The most promising uses of the method are for investigating facial features for conditions in which we do not have a priori knowledge of the features or of their intensities or we do not want to bias results by providing features to the participants. This, for example, would be most useful for research in racial differences in facial expression of emotion. Also, current work includes research on the shape of the internal representation of facial features, e.g., whether the face space is linear or if it contains significant warping or scaling (see Poirier & Faubert, [in process](#)).

Acknowledgements

This research was supported by an NSERC-Essilor industrial research chair and an NSERC discovery fund awarded to Jocelyn Faubert. Portions of this paper were presented at Vision Sciences Society 2010 in Naples, Florida.

Commercial relationships: none.

Corresponding author: Frédéric J.A.M. Poirier.

Email: jamfpo@yahoo.com.

Address: Université de Montréal, Département D'Optométrie, Montréal, Québec, Canada.

References

- Ambadar, S., Schooler, J., & Cohn, J. (2005). Deciphering the enigmatic face: the importance of facial dynamics in interpreting subtle facial expressions. *Psychological Science*, *16*(5), 403–410.
- Anderson, N. D., & Wilson, H. R. (2005). The nature of synthetic face adaptation. *Vision Research*, *45*(14), 1815–1828. doi: 10.1016/j.visres.2005.01.012.
- Barton, J. J., Keenan, J. P., & Bass, T. (2001). Discrimination of spatial relations and features in faces: Effects of inversion and viewing duration. *British Journal of Psychology*, *92*, 527–549.
- Bassili, J. N. (1978). Facial motion in the perception of faces and of emotional expressions. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 373–379.
- Bassili, J. N. (1979). Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face. *Journal of Personality and Social Psychology*, *37*, 2049–2058.
- Benson, P. J., & Perrett, D. I. (1994). Visual processing of facial distinctiveness. *Perception*, *23*, 75–93.
- Benton, C. P., Etchells, P. J., Porter, G., Clark, A. P., Penton-Voak, I. S., & Nikolov, S. G. (2007). Turning the other cheek: the viewpoint dependence of facial expression after-effects. *Proceedings of the Royal Society: B*, *274*, 2131–2137, doi: 10.1098/rspb.2007.0473.
- Biele, C., & Grabowska, A. (2006). Sex differences in perception of emotion intensity in dynamic and static facial expressions. *Experimental Brain Research*, *171*, 1–6, doi: 10.1007/s00221-005-0254-0.
- Blanz, V., O'Toole, A. J., Vetter, T., & Wild, H. A. (2000). On the other side of the mean: The perception of dissimilarity in human faces. *Perception*, *29*, 885–891.
- Boucher, J. D., & Ekman, P. (1975). Facial areas and emotional information. *The Journal of Communication*, *25*(2), 21–29.
- Brown, E., & Perrett, D. I. (1993). What gives a face its gender. *Perception*, *22*, 829–840.
- Bruce, V., Hanna, E., Dench, N., Healey, P., & Burton, M. (1992). The importance of mass in line drawings of faces. *Applied Cognitive Psychology*, *6*(7), 619–628.
- Brunswik, E., & Reiter, L. (1937). Eindruckscharaktere schematisierter gesichter. *Zeitschrift fuer Psychologie*, *142*, 67–134.
- Burton, A. M., Bruce, V., & Dench, N. (1993). What's

- the difference between men and women? Evidence from facial measurement. *Perception*, 22(2), 153–176.
- Butler, S., Blais, C., Gosselin, F., Bub, D., & Fiset, D. (2010). Recognizing famous people. *Attention, Perception, & Psychophysics*, 72, 1444–1449, doi: 10.3758/APP.72.6.1444.
- Caldara, R., Schyns, P. G., Mayer, E., Smith, M. L., Gosselin, F., & Rossion, B. (2005). Does prosopagnosia take the eyes out of face representations? Evidence for a defect in representing diagnostic facial information following brain damage. *Journal of Cognitive Neuroscience*, 17, 1652–1666, doi: 10.1162/089892905774597254.
- Campbell, R., Benson, P. J., Wallace, S. B., Doesbergh, S., & Coleman, M. (1999). More about brows: How poses that change brow position affect perceptions of gender. *Perception*, 28(4), 489–504.
- Chen, H., Russell, R., Nakayama, K., & Livingstone, M. (2010). Crossing the ‘uncanny valley’: adaptation to cartoon faces can influence perception of human faces. *Perception*, 39(3), 378–386.
- Cheung, O. S., Richler, J. J., Palmeri, T. J., & Gauthier, I. (2008). Revisiting the role of spatial frequencies in the holistic processing of faces. *Journal of experimental psychology. Human perception and performance*, 34(6), 1327–1336.
- Cook, R., Matei, M., & Johnston, A. (2011). Exploring expression space: Adaptation to orthogonal and anti-expressions. *Journal of Vision*, 11(4):2, 1–9, <http://www.journalofvision.org/content/11/4/2>, doi: 10.1167/11.4.2. [PubMed] [Article].
- Cunningham, D. W., & Wallraven, C. (2009a). Dynamic information for the recognition of conversational expressions. *Journal of Vision*, 9(13):7, 1–17, <http://journalofvision.org/9/13/7>, doi: 10.1167/9.13.7. [PubMed] [Article].
- Cunningham, D. W., & Wallraven, C. (2009b). The interaction between motion and form in expression recognition. In Bodenheimer B. & O’Sullivan C. (Eds.), *Proceedings of the 6th Symposium on Applied Perception in Graphics and Visualization (APGV2009)*. New York, NY: ACM.
- Dakin, S. C., & Omigie, D. (2009). Psychophysical evidence for a non-linear representation of facial identity. *Vision Research*, 49(18), 2285–2296, doi: 10.1016/j.visres.2009.06.016.
- Dupuis-Roy, N., Fortin, I., Fiset, D., & Gosselin, F. (2009). Uncovering gender discrimination cues in a realistic setting. *Journal of Vision*, 9(2):10, 1–8, <http://journalofvision.org/9/2/10/>, doi: 10.1167/9.2.10. [PubMed] [Article].
- Ehrlich, S. M., Schiano, D. J., & Sheridan, K. (2000). Communicating facial affect: it’s not the realism, it’s the motion. *Proceedings of ACM CHI 2000 conference on human factors in computing systems*. New York, NY: ACM.
- Ekman, P., & Friesen, W. V. (1975). *Unmasking the face. A guide to recognizing emotions from facial clues*. New Jersey: Prentice-Hall.
- Ekman, P., & Friesen, W. V. (1976). Measuring facial movement. *Environmental Psychology and Nonverbal Behavior*, 1(1), 56–75.
- Ekman, P., & Friesen, W. V. (1978). *The Facial Action Coding System*. Palo Alto: Consulting Psychologists’ Press.
- Ellamil, M., Susskind, J. M., & Anderson, A. K. (2008). Examinations of identity invariance in facial expression adaptation. *Cognitive, Affective, and Behavioral Neuroscience*, 8(3), 273–281.
- Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1979). Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. *Perception*, 8, 431–439, doi: 10.1068/p080431.
- Farah, M. J., Wilson, K. D., Drain, H. M., & Tanaka, J. R. (1995). The inverted face inversion effect in prosopagnosia: Evidence for mandatory, face-specific perceptual mechanisms. *Vision Research*, 35(14), 2089–2093.
- Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is “special” about face perception? *Psychology Review*, 105(3), 482–498.
- Fox, C. J., & Barton, J. S. S. (2007). What is adapted in face adaptation? The neural representations of expression in the human visual system. *Brain Research*, 1127, 80–89.
- Garrod, O., Yu, H., Breidt, M., Curio, C., & Schyns, P. (2010). Reverse correlation in temporal FACS space reveals diagnostic information during dynamic emotional expression classification. *Journal of Vision*, 10(7): 700, <http://www.journalofvision.org/content/10/7/700>, doi: 10.1167/10.7.700. [Abstract].
- Gauthier, I., & Bukach, C. M. (2007). Should we reject the expertise hypothesis? *Cognition*, 103, 322–330.
- Gauthier, I., & Tarr, M. J. (2002). Unraveling mechanisms for expert object recognition: Bridging brain activity and behavior. *Journal of Experimental Psychology Human Perception and Performance*, 28, 431–446.
- Goffaux, V., & Rossion, B. (2006). Faces are “Spatial”—Holistic face perception is supported by low spatial frequencies. *Journal of Experimental Psy-*

- chology: *Human Perception and Performance*, 32(4), 1023–1039.
- Goffaux, V., & Rossion, B. (2007). Face inversion disproportionately impairs the perception of vertical but not horizontal relations between features. *Journal of Experimental Psychology: Human Perception and Performance*, 33, 995–1002.
- Goren, D., & Wilson, H. R. (2006). Quantifying facial expression recognition across viewing conditions. *Vision Research*, 46(8-9), 1253–1262.
- Gosselin, F., & Schyns, P. G. (2001). Bubbles: a technique to reveal the use of information in recognition tasks. *Vision Research*, 41(17), 2261–2271, doi: 10.1016/S0042-6989(01)00097-9.
- Gota, K., & Miyamoto, M. (2000). Emotion recognition. Facial components associated with various emotions. *Japanese Journal of Psychology*, 71(3), 211–218.
- Harwood, N. K., Hall, L. J., & Shinkfield, A. J. (1999). Recognition of facial emotional expressions from moving and static displays by individuals with mental retardation. *American Journal of Mental Retardation*, 104(3), 270–278.
- Hsu, S. M., & Young, A. W. (2004). Adaptation effects in facial expression recognition. *Visual Cognition*, 11, 871–899.
- Izard, C. E. (1971). *The face of emotion*. New York: Appleton Century Crofts.
- Izard, C. E. (1994). Innate and universal facial expressions: Evidence from developmental and cross-cultural research. *Psychological Bulletin*, 2, 288–299.
- Johnston, R. A., Milne, A. B., Williams, C., & Hosie, J. (1997). Do distinctive faces come from outer space? An investigation of the status of a multidimensional face-space. *Visual Cognition*, 4(1), 59–67, doi: 10.1080/713756748.
- Jones, B. C., Feinberg, D. R., Bestelmeyer, P. E. G., DeBruine, L. M., & Little, A. C. (2010). Adaptation to different mouth shapes influences visual perception of ambiguous lip speech. *Psychonomic Bulletin & Review*, 17(4), 522–528, doi: 10.3758/PBR.17.4.522.
- Kätsyri, J., & Sams, M. (2008). The effect of dynamics on identifying basic emotions from synthetic and natural faces. *International Journal of Human-Computer Studies*, 66, 233–242.
- Kilts, C. D., Egan, G., Gideon, D. A., Ely, T. D., & Hoffman, J. M. (2003). Dissociable neural pathways are involved in the recognition of emotion in static and dynamic facial expressions. *NeuroImage*, 18(1), 156–168.
- Knappmeyer, B., Thornton, I. M., & Bülthoff, H. H. (2003). The use of facial motion and facial form during the processing of identity. *Vision Research*, 43(18), 1921–1936.
- Konar, Y., Bennett, P. J., & Sekuler, A. B. (2010). Holistic processing is not correlated with face-identification accuracy. *Psychological Science*, 21(1), 38–43, doi: 10.1177/0956797609356508.
- Kostov, V., Yanagisawa, H., Johansson, M., & Fukuda, S. (2001). Method for face-emotion retrieval using a cartoon emotional expression approach. *JSME International Journal Series C*, 44(2), 515–526, doi: 10.1299/jsmec.44.515.
- LaBar, K. S., Crupain, M. J., Voyvodic, J. T., & McCarthy, G. (2003). Dynamic perception of facial affect and identity in the human brain. *Cerebral Cortex*, 13(10), 1023–1033.
- Lander, K. (2005). Why are moving faces easier to recognize? *Visual Cognition*, 12(3), 429–442.
- Leder, H., & Carbon, C. C. (2004). Part-to-whole effects and configural processing in faces. *Psychology and Science*, 46, 531–543.
- Lee, K., Byatt, G., & Rhodes, G. (2000). Caricature effects, distinctiveness, and identification: Testing the face-space framework. *Psychological Science*, 11, 379–385.
- Leopold, D. A., O’Toole, A. J., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nature Neuroscience*, 4(1), 89–94.
- Loffler, G., Gordon, G. E., Wilkinson, F., Goren, D., & Wilson, H. R. (2005). Configural masking of faces: Evidence for high-level interactions in face perception. *Vision Research*, 45(17), 2287–2297, doi: 10.1016/j.visres.2005.02.009.
- Loffler, G., Wilkinson, F., Yourganov, G., & Wilson, H. R. (2004). Effect of facial geometry on the fMRI signal in the fusiform face area. *Journal of Vision*, 4(8): 136a, <http://journalofvision.org/4/8/136/>, doi: 10.1167/4.8.136. [Abstract].
- Loffler, G., Yourganov, G., Wilkinson, F., & Wilson, H. R. (2005). fMRI evidence for the neural representation of faces. *Nature Neuroscience*, 8(10), 1386–1390.
- Malcom, G. L., Leung, C., & Barton, J. (2005). Regional variation in the inversion effect for faces: Differential effects for feature shape, feature configuration and external contour. *Perception*, 34, 1221–1231.
- Mangini, M. C., & Biederman, I. (2004). Making the ineffable explicit: estimating the information em-

- ployed for face classifications. *Cognitive Science*, 28, 209–226, doi: 10.1016/j.cogsci.2003.11.004.
- Maurer, D., Le Grand, R., & Mondlock, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, 6(6), 255–260.
- Mauro, R., & Kubovy, M. (1992). Caricature and face recognition. *Memory and Cognition*, 20, 433–440.
- McKone, E., Kanwisher, N., & Duchaine, B. C. (2007). Can generic expertise explain special processing for faces? *Trends in Cognitive Sciences*, 11(1), 8–15.
- O’Toole, A. J., Roark, D. A., & Abdi, H. (2002). Recognizing moving faces: A psychological and neural synthesis. *Trends in Cognitive Sciences*, 6(6), 261–266, doi: 10.1016/S1364-6613(02)01908-3.
- Plutchik, R. (1980). *Emotions: A psychoevolutionary synthesis*. New York: Harper & Row.
- Plutchik, R. (2001). The nature of emotions. *American Scientist*, 89, 344–350.
- Poirier, F. J. A. M., & Faubert, J. (*in process*). The face space for happy/sad and gender is not linear: Empirical mapping of the perceptual representation.
- Rhodes, G. (1996). *Superportraits: Caricatures and recognition*. Hove: The Psychology Press.
- Rhodes, G., Brennan, S., & Carey, S. (1987). Identification and ratings of caricatures: implications for mental representations of faces. *Cognitive Psychology*, 19, 473–497.
- Rhodes, G., & Jeffery, L. (2006). Adaptive norm-based coding of facial identity. *Vision Research*, 46, 2977–2987.
- Rhodes, G., Maloney, L. T., Turner, J., & Ewing, L. (2007). Adaptive face coding and discrimination around the average face. *Vision Research*, 47, 974–989.
- Rhodes, G., & Tremewan, T. (1994). Understanding face recognition: Caricature effects, inversion, and the homogeneity problem. *Visual Cognition*, 1, 275–311.
- Rhodes, G., & Tremewan, T. (1996). Averageness, exaggeration, and facial attractiveness. *Psychological Science*, 7, 105–110.
- Richler, J. J., Bukach, C. M., & Gauthier, I. (2009). Context influences holistic processing of non-face objects in the composite task. *Attention, Perception and Psychophysics*, 71, 530–540.
- Richler, J. J., Gauthier, I., Wenger, M. J., & Palmeri, T. J. (2008). Holistic processing of faces: Perceptual and decisional components. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 328–342.
- Richler, J. J., Tanaka, J. W., Brown, D. D., & Gauthier, I. (2008). Why does selective attention to parts fail in face processing? *Journal of Experimental Psychology: Learning, Memory and Cognition*, 34, 1356–1368.
- Roark, D. A., Barrett, S. E., Spence, M. J., Hervé, A., & O’Toole, A. J. (2003). Psychological and neural perspectives on the role of motion in face recognition. *Behavioral and Cognitive Neuroscience Reviews*, 2(1), 15–46.
- Robbins, R., & McKone, E. (2007). No face-like processing for objects-of-expertise in three behavioural tasks. *Cognition*, 103(1), 34–79, doi: 10.1016/j.cognition.2006.02.008.
- Rossion, B. (2008). Picture-plane inversion leads to qualitative changes of face perception. *Acta Psychologica*, 128, 274–289.
- Rossion, B. (2009). Distinguishing the cause and consequence of face inversion: The perceptual field hypothesis. *Acta Psychologica*, 132, 300–312.
- Roy, S., Roy, C., Fortin, I., Ethier-Majcher, C., Belin, P., & Gosselin, F. (2007). A dynamic facial expression database. *Journal of Vision*, 7(9): 944, <http://www.journalofvision.org/content/7/9/944>, doi: 10.1167/7.9.944. [Abstract].
- Russell, R. (2003). Sex, beauty, and the relative luminance of facial features. *Perception*, 32(9), 1093–1107.
- Russell, R. (2005). Face pigmentation and sex classification. *Journal of Vision*, 5(8): 983, 983a, <http://journalofvision.org/5/8/983/>, doi: 10.1167/5.8.983. [Abstract].
- Russell, R. (2009). A sex difference in facial contrast and its exaggeration by cosmetics. *Perception*, 38(8), 1211–1219.
- Rutherford, M. D., Chattha, H. M., & Krysko, K. M. (2008). The use of aftereffects in the study of relationships among emotion categories. *Journal of Experimental Psychology: Human Perception and Performance*, 34(1), 27–40.
- Sato, W., Kochiyama, T., Yoshikawa, S., Naito, E., & Matsumura, M. (2004). Enhanced neural activity in response to dynamic facial expressions of emotion: an fMRI study. *Cognitive Brain Research*, 20(1), 81–91, doi: 10.1016/j.cogbrainres.2004.01.008.
- Schyns, P. G., Bonnar, L., & Gosselin, F. (2002). Show me the features! Understanding recognition from the use of visual information. *Psychological Science*, 13, 402–409, doi: 10.1111/1467-9280.00472.
- Sekuler, A. B., Gaspar, C. M., Gold, J. M., & Bennett, P. J. (2004). Inversion leads to quantitative, not qualitative, changes in face processing. *Current*

- Biology*, 14, 391–396, doi: 10.1016/j.cub.2004.02.028.
- Sekunova, A., & Barton, J. (2008). The effects of face inversion on the perception of long-range and local spatial relations in eye and mouth configurations. *Journal of Experimental Psychology: Human Perception and Performance*, 34(5), 1129–1135.
- Sigala, N., & Logothetis, N. K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature*, 415(17), 318–320.
- Simon, D., Craig, K. D., Gosselin, F., Belin, P., & Rainville, P. (2008). Recognition and discrimination of prototypical dynamic expressions of pain and emotions. *Pain*, 135, 55–64.
- Skinner, A. L., & Benton, C. P. (2010). Anti-expression aftereffects reveal prototype-referenced coding of facial expressions. *Psychological Science*, 21, 1248–1253.
- Smith, M. L., Cottrell, G. W., Gosselin, F., & Schyns, P. G. (2005). Transmitting and decoding facial expressions. *Psychological Science*, 16(3), 184–189.
- Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology*, 46A, 225–245.
- Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory and Cognition*, 25, 583–592.
- Taschereau-Dumouchel, V., Rossion, B., Schyns, P. G., & Gosselin, F. (2010). Interattribute distances do not represent the identity of real-world faces. *Frontiers in Psychology*, 1(159), 1–10, doi: 10.3389/fpsyg.2010.00159.
- Tyler, C. W., & Chen, C.-C. (2006). Spatial summation of face information. *Journal of Vision*, 6(10):11, 1117–1125, <http://www.journalofvision.org/content/6/10/11>, doi: 10.1167/6.10.11. [PubMed] [Article].
- Valentine, T. (1988). Upside-down faces: a review of the effect of inversion upon face recognition. *British Journal of Psychology*, 79(4), 471–491.
- Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion and race in face recognition. *Quarterly Journal of Experimental Psychology*, 43A, 161–204.
- Valentine, T. (2001). Face-Space Models of Face Recognition. In Wenger, M. J. & Townsend, J. T. (Eds.) *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges*. Hillsdale, New Jersey: Lawrence Erlbaum Associates Inc.
- Valentine, T., & Bruce, V. (1986). The effects of distinctiveness in recognising and classifying faces. *Perception*, 15, 525–535.
- Valentine, T., Darling, S., & Donnelly, M. (2004). Why are average faces attractive? The effect of view and averageness on the attractiveness of female faces. *Psychonomic Bulletin & Review*, 11(3), 482–487.
- Van Belle, G., De Graef, P., Verfaillie, K., Rossion, B., & Lefèvre, P. (2010). Face inversion impairs holistic perception: Evidence from gaze-contingent stimulation. *Journal of Vision*, 10(5):10, 1–13, <http://www.journalofvision.org/content/10/5/10>, doi:10.1167/10.5.10. [PubMed] [Article].
- Webster, M. A., Kaping, D., Mizokami, Y., & Duhamel, P. (2004). Adaptation to natural facial categories. *Nature*, 428, 557–561.
- Webster, M. A., & MacLin, O. H. (1999). Figural aftereffects in the perception of faces. *Psychonomic Bulletin & Review*, 6, 647–653.
- Wenger, M. J., & Ingvalson, E. M. (2002). A decisional component of holistic encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 872–892.
- Wilson, H. R., Loffler, G., & Wilkinson, F. (2002). Synthetic faces, face cubes, and the geometry of face space. *Vision Research*, 42, 2909–2923.
- Yamaguchi, M. K., Hirukawa, T., & Kanazawa, S. (1995). Judgment of gender through facial parts. *Perception*, 24(5), 563–575.
- Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, 81, 141–145.
- Young, A. W., Hay, D. C., McWeeny, K. H., Flude, B. M., & Ellis, A. W. (1985). Matching familiar and unfamiliar faces on internal and external features. *Perception*, 14, 737–746, doi: 10.1068/p140737.
- Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception*, 16(6), 747–759.
- Yovel, G., & Kanwisher, N. (2005). The neural basis of the behavioural face-inversion effect. *Current Biology*, 15, 2256–2262.