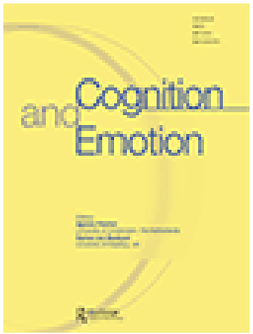

Routledge
Taylor & Francis Group

Cognition and Emotion

ISSN: 0269-9931 (Print) 1464-0600 (Online) Journal homepage: http://www.tandfonline.com/loi/pcem20

Development and validation of an Argentine set of facial expressions of emotion

Marcelo Vaiman, Mónica Anna Wagner, Estefanía Caicedo & Germán Leandro Pereno

To cite this article: Marcelo Vaiman, Mónica Anna Wagner, Estefanía Caicedo & Germán Leandro Pereno (2015): Development and validation of an Argentine set of facial expressions of emotion, Cognition and Emotion, DOI: 10.1080/02699931.2015.1098590

To link to this article: http://dx.doi.org/10.1080/02699931.2015.1098590

COGNITION AND EMOTION, 2015
http://dx.doi.org/10.1080/02699931.2015.1098590

Routledge
Taylor & Francis Group

Development and validation of an Argentine set of facial expressions of emotion

Marcelo Vaiman, Mónica Anna Wagner, Estefanía Caicedo and Germán Leandro Pereno

Cognitive Psychology Laboratory, Faculty of Psychology, National University of Córdoba, Cordoba, Argentina

ABSTRACT

Pictures of facial expressions of emotion are used in a wide range of experiments. The last decade has seen an increase in the number of studies presenting local sets of emotion stimuli. However, only a few existing sets contain pictures of Latin Americans, despite the growing attention emotion research is receiving in this region. Here we present the development and validation of the Universidad Nacional de Cordoba, Expresiones de Emociones Faciales (UNCEEF), a Facial Action Coding System (FACS)-verified set of pictures of Argentineans expressing the six basic emotions, plus neutral expressions. FACS scores, recognition rates, Hu scores, and discrimination indices are reported. Evidence of convergent validity was obtained using the Pictures of Facial Affect in an Argentine sample. However, recognition accuracy was greater for UNCEEF. The importance of local sets of emotion pictures is discussed.

ARTICLE HISTORY
Received 28 November 2014
Revised 6 September 2015
Accepted 17 September 2015

KEYWORDS
Facial expressions of emotion; stimulus set; FACS; validation; Argentina

Emotional facial expression stimulus sets

Facial expressions of emotion are used as stimuli in studies from a wide range of fields, such as cognitive psychology, artificial intelligence, and neuroscience. Since each field has different needs with regard to their stimuli, these can vary on multiple dimensions, for example: whether they are static (pictures) or dynamic (videos), eye gaze direction, head orientation, how much of the face and/or body is pictured, or characteristics inherent to the models expressing the emotions, such as age, sex, and ethnicity. The last one is relevant in light of the discussion on the so-called in-group advantage for emotion recognition (e.g., Beaupré & Hess, 2005; Biehl et al., 1997; Elfenbein & Ambady, 2002; Jack, Garrod, Yu, Caldara, & Schyns, 2012; Matsumoto, 2002) according to which individuals more accurately recognise emotions when expressed by members of their own cultural group versus members of another group.

Since the last decade the number of studies presenting local sets of pictures of facial expressions of emotion has increased (for an overview of the existing sets, see Table 1). However, to our knowledge, only a few current sets contain pictures of Latin Americans (e.g., Gur et al., 2002; Tottenham et al., 2009). This gap is particularly notable considering that a recent bibliometric study (Jeanneret, Rego, Oña, Vaiman, & Pereno, 2015) found that 4% of articles published on facial expressions of emotion in the last five years were from Latin American countries. Stimulus sets that do include models from this population usually refer to them with the general term "Hispanics". Although Latin American countries share a language and geographical area, like many large regions, each community has distinct physiognomic features, which are a product of different sociocultural and genetic influences (del Sol, 2006). The racial heritage of Argentina, for example, is the result of a large wave of immigration, mostly from Italy and Spain, which joined the native indigenous peoples who inhabited the territory (Magrassi, 1999; for genetic evidence, see Corach et al., 2010).

To date, emotion studies carried out in Latin America have mainly used the Pictures of Facial Affect (POFA; Ekman & Friesen, 1976), a set of pictures of North Americans expressing the six basic emotions. However, a

CONTACT Germán Leandro Pereno gpereno@psyche.unc.edu.ar

Table 1. Examples of the existing face databases, their features, and hit rates

Name	Authors (year)	Ethnicities models	Emotions evaluated (hit rates)
Pictures of Facial Affect (POFA)	Ekman and Friesen (1976)	Caucasian	Happiness (.99), sadness (.89), fear (.88), anger (.89), surprise (.92), and disgust (.92)
Japanese and Caucasian Facial Expressions of Emotion/Japanese and Caucasian Neutral Faces (JACFEE/ JACNeuF)	Matsumoto and Ekman (1988)	Caucasian and Japanese	Happiness (.98), sadness (.88), fear (.67), anger (.80), surprise (.91), disgust (.76), and contempt (.79)
Karolinska Directed Emotional Faces (KDEF)	Lundqvist, Flykt, and Öhman (1998); validation study: Goeleven et al. (2008)	Caucasian	Happiness (.93), sadness (.77), fear (.43), anger (.79), surprise (.77), disgust (.72), and neutral (.63)
NimStim Set of Facial Expressions	Tottenham et al. (2009)	African-, Asian-, European-, and Latin American	Happiness (.99), sadness (.59), fear (.74), anger (.96), surprise (.86), disgust (.93), calm (.81), and neutral (.86)
University of California, Davis, Set of Emotion Expressions (UCDSEE)	Tracy et al. (2009)	Caucasian and African	Happiness (.94), sadness (.80), fear (.51), anger (.74), surprise (.92), disgust (.81), embarrassment (.61), pride (.89), and shame (.47)
Radboud Faces Database	Langner et al. (2010)	Dutch Caucasian	Happiness (.98), sadness (.80), fear (.83), anger (.85), surprise (.90), disgust (.81), contempt (.53), and neutral (.84)
FACES	Ebner, Riediger, and Lindenberger (2010)	Caucasian	Happiness (.96), sadness (.73), fear (.81), anger (.81), disgust (.68), and neutral (.87)
Amsterdam Dynamic Facial Expression Set (ADFES)	van der Schalk et al. (2011)	North-European and Mediterranean	Happiness (.91), sadness (.82), fear (.84), anger (.88), surprise (.89), disgust (.86), contempt (.68), embarrassment (.74), and pride (.69)
Penn Emotion Recognition Task	Gur et al. (2002)	Caucasian, African American, Asian, and Hispanic	Happiness (.96), sadness (.74), fear (.63), anger (.53), disgust (.56), and neutral (.76)

study (Vaiman, Caicedo, & Pereno, 2011) in which this set was administered in Argentina yielded recognition rates lower than those found in the USA in Ekman and Friesen's (1976) original study. Thus, a local set of emotion pictures is necessary. Here we present the development and validation of a set of pictures of Argentine models expressing the basic emotions, which we refer to as the Universidad Nacional de Cordoba, Expresiones de Emociones Faciales (UNCEEF).

Study 1. Stimulus development

The first aim of this project was to obtain a set of pictures of emotions as expressed by Argentineans. The target expressions were the prototypes and major variants of the six basic emotions as defined by Ekman, Friesen, and Hager (2002b). To this end, the Facial Action Coding System (FACS; Ekman & Friesen, 1978; Ekman, Friesen, & Hager, 2002a) was used. FACS is an anatomically based system for coding facial behaviour that has been used as an objective method to create standardised stimuli for emotion studies. However, it has been suggested that posed emotions may not be identical to authentic expressions (e.g., Naab & Russell, 2007; Russell, 1994). To address these concerns, two strategies were used to elicit emotions: emotion induction and FACS-guided posing.

Method

Pre-session

Male and female individuals were recruited to an "audition" where their ability to produce facial expressions of emotion was evaluated. Those who demonstrated a greater ability to express emotions were invited to participate in the photo shoot. Face models were informed about the general aim of the project, as well as the specific procedure, then signed the informed consent document permitting the use of their pictures for research purposes.

Participants

Fourteen models (8 males and 6 females) between the ages of 18 and 44 ($M = 25.53$, $SD = 8.72$) participated in the photo shoot. Models were all Argentine citizens and residents with mainly Southern European (Spanish and Italian), but also Western (French) and Eastern (Polish and Russian) European and West Asian (Lebanese and Syrian) ancestry, reflecting the ethnic diversity of Argentina.

Procedure and materials

Photo shoot. The models wore plain white T-shirts and stood in front of a neutral blue background. Models were asked to pull their hair back, if it covered part of their face. All the photographs were taken head on and from the shoulders up. High-quality digital photographs were taken using a 10-Mpx Nikon D3000 camera equipped with flash (1500 WS). All the pictures were in colour mode sRGB.

Instruction. The models were told that they would be photographed several times, expressing different emotions. They were instructed to look straight ahead into the camera lens. Throughout the session, the researchers continuously coached the models, giving them instructions and feedback on their expressions. Facial expressions of emotion were elicited in two stages. The first phase aimed at inducing each emotion by triggering its subjective experience. Models were instructed to display each facial expression as intensely as possible, while maintaining its natural look. To this end, they were asked to think of and re-experience a situation in their personal past in which they had felt the emotion and to display it. To warm up, the session started with neutral facial expressions followed by anger, sadness, disgust, happiness, anger, and, finally, fear.

In the second phase the models received instructions on how to express each emotion based on the directed facial action task (Ekman, 2007; Levenson, Carstensen, Friesen, & Ekman, 1991). For each picture, a researcher stood in front of the model, coaching them until the target expression was obtained. Models were offered a mirror in order to help them target the desired action units (AUs). Target expressions consisted in prototypical and major variants of each emotion at different intensities, as specified in Ekman et al. (2002b; Ekman, 2007; see Table 2 for a summary). On average, 30–40 pictures were taken per model.

Table 2. Targeted AU configurations for each emotion

Emotion	FACS scores
Happiness	6+12 + [25 AND/OR 26]
Sadness	1 + 4 + 15 + [17]
Fear	1 + 2 + 4 + 5 + [20] + [25 AND/OR 26/27]
Anger	4 + 5 + 7 + [17] + 23/24
Disgust	9/10 + [16/17] + [25 AND/OR 26]
Surprise	1 + 2 + 5AB + [25 AND/OR 26/27]

Note: Targeted configurations were based on the prototypical expressions and their major variants presented in Ekman et al. (2002b) and Ekman (2007). Letters refer to AU intensity.

Validation of the UNCEEF database

Study 2. FACS verification

In the first study, pictures of facial expressions of emotion were obtained by inducing and posing prototypical and major variants of AU configurations of each emotion. The aim of the present study was to obtain content-based evidence of validity (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). In order to achieve this, the best pictures from the photo shoot were FACS scored by experts and only those containing the target AUs were retained.

Method

The pictures from the photo shoot were pre-screened by the authors to select those that most corresponded to the targeted expressions in terms of AUs (see Table 2). The selected pictures were then coded by external certified FACS coders[1] to insure objectivity. Of all the pictures obtained from the photo shoot, 359 were submitted for codification.

One certified FACS coder undertook the full FACS coding of all 359 pictures, including intensity ratings. Then two more certified FACS experts randomly coded 10% of the photographs in order to assess the inter-rater reliability according to the formula provided by Ekman et al. (2002b). An inter-rater reliability coefficient of .87 was obtained, exceeding the standard of .70. Finally, the coders discussed the scores until reaching 100% agreement. Only those AUs completely agreed upon were included in the final FACS scores for the pictures.

Next, the pictures which best represented the prototypes and their major variants of each emotion according to Ekman et al. (2002b; Ekman, 2007) were retained. The selection criteria were that the pictures: (1) include AUs essential to a prototype or major variant of the target emotion and (2) exclude AUs not relevant to the target emotion, especially AUs essential to other emotions (Ekman, 2007; Rosenberg & Ekman, 1995). An attempt was made to include different versions and intensities of each emotion.

Results

Of the 359 original pictures sent for FACS coding, 100 were selected based on the concordance

between their FACS scores and the target expression of that emotion. Of the 100 FACS-verified pictures, 16 were of happiness, 15 sadness, 12 fear, 15 anger, 15 surprise, 16 disgust, and 15 neutral. Furthermore, 20 were naturally posed and 80 were obtained by the directed facial action task. At least one picture of each of the initial 14 models remained.

Discussion

For some emotions, it was difficult to find "pure" expressions. For example, happy expressions often included AU 7, AU 5 often exceeded the maximum intensity (B) indicated for surprise, and fear expressions often lacked AU 2. In these cases, care was taken that the extra AUs were not critical to any other emotions. Furthermore, as Ekman et al. (2002b) note, exhaustive empirical evidence for the AUs corresponding to each emotion is lacking, particularly for the distinction between surprise and fear and the lower face AUs for sadness. Therefore, in the following study we put the pictures to the test by asking participants to judge what emotion each picture expresses. This way, the set could have an additional source of validity.

Study 3. Hit rates and discrimination indices

Emotion researchers often need stimuli of varying levels of difficulty. Therefore, while all pictures should be reliably recognised, we also wished to include diversity in our set, with pictures that are more or less difficult to decode. Moreover, pictures of facial expressions of emotion are often used for diagnostic purposes where emotion recognition is evaluated. Thus, an index of the ability of a picture to discriminate between the people with the most and least ability to recognise emotions is also valuable. To this end, we also evaluated the discriminative power of the pictures by calculating their discrimination indices. The aim of the present study was therefore to select pictures that were relatively highly recognisable, but that also demonstrated high discriminative power.

Method

Participants

The non-probabilistic accidental sample consisted of 466 students (77% female and 23% male) from the Faculty of Psychology at the National University of Cordoba, Argentina, between the ages of 16 and 39 ($M = 20.29$, $SD = 4.33$).

Measure

The 100 pictures selected in the previous study were administered. In order to improve the quality of the pictures and produce greater uniformity, the pictures were edited (e.g., retouched and colour-matched) using the GNU Image Manipulation Program (GIMP). In addition, the pictures were resized to 2126×2244 pixel resolution and saved in JPEG format.

The set was divided into two sub-sets, presented with a break in between, to prevent fatigue. All of the pictures of each encoder were presented in the same sub-set and each sub-set was equivalent in the total number of pictures, number of pictures per emotion, and number of male and female encoders (or as close as possible). Within each sub-set, the pictures were presented in random order, identical for all of the participants. The pictures were shown in a PowerPoint presentation. Each picture was presented for five seconds, followed by a blank screen during three seconds. The pictures were numbered according to their order of appearance and this number was presented both visually and auditorily during the same slide as the picture. The stimuli were projected onto a large 300 cm × 220 cm screen.

Procedure

The pictures were administered collectively, although participants were instructed to complete the task individually, responding on a sheet of paper. For each of the 100 pictures, participants were instructed to decide which emotion, if any, they thought was being expressed, choosing the emotion that best matched the emotion expressed by the person in the picture. They were given the response options of anger, disgust, fear, happiness, sadness, surprise, no emotion, and other,[2] which included a blank space where participants could respond in an open ended manner. The "other" and "no emotion" options were included since concerns have been raised about the

[1]The FACS coding was carried out by the Zurich Interaction and Expression Lab, Department of Psychology, University of Zurich, Switzerland.
[2]Options were presented in Spanish as enojo, asco, miedo, alegría, tristeza, sorpresa, ninguna, and otro, respectively. These labels were chosen following prior studies finding them optimal (Vaiman et al., 2011).

Table 3. Amount of pictures overall, by sex of the encoder, and by encoder for each emotion

Emotion	Overall	Sex of encoder		Encoder													
		Male	Female	C	D	E	F	G	H	I	K	L	N	O	P	S	U
Happiness	9	5	4	1	1	0	0	1	0	1	0	1	1	1	0	1	1
Sadness	8	5	3	1	0	1	1	0	0	1	1	0	0	2	0	1	0
Fear	6	2	4	4	0	0	0	1	0	0	0	0	0	1	0	0	0
Anger	9	6	3	1	1	1	0	1	0	1	1	0	0	1	1	1	0
Surprise	9	6	3	1	0	1	1	0	1	1	0	0	0	1	1	1	1
Disgust	9	6	3	0	1	0	1	1	1	0	1	0	1	2	0	0	1
Neutral	10	6	4	1	1	1	1	1	0	1	1	0	0	0	1	1	1
Total	60	36	24	9	4	4	4	5	2	5	4	1	2	8	3	5	4

traditional forced-choice response method (e.g., Frank & Stennett, 2001; Russell, 1994).

Results

Recognition rates were obtained by first designating responses as correct or incorrect. "Other" responses considered synonyms of the target emotion (e.g., joy for happiness) were counted as correct responses. Other responses that did not include an alternative or were not considered synonyms (e.g., doubt for any emotion), responses left blank, along with non-target emotion responses, were all considered incorrect. Recognition rates were then calculated as the proportion of correct responses or "hits" for each photograph. The data of 30 participants were excluded from the analyses for presenting recognition rates $<= 3.29$ SDs overall or in more than three emotions (Tabachnick & Fidell, 2001).

Since hit rates do not take into account judge response bias, we also calculated unbiased hit rates (*Hu* scores; Wagner, 1993) which are the ratio of the number of hits and the number of stimuli of the target emotion adjusted for the number of times participants erroneously chose the same label for other displays (false alarm rates). Hu scores were calculated per model (hereafter "encoder"), for each photograph (Goeleven, De Raedt, Leyman, & Verschuere, 2008).

While high hit rates are desirable, pictures with high discriminative power are also valuable. To this end, we obtained the discrimination index of each picture, a measure of how well a picture is able to distinguish between those participants who obtained the highest scores overall and those with the lowest. This index is calculated as the difference between the hit rates of the two groups, that is, the participants located in the third quartile and those in the first (Ebel, 1965).

Moreover, these indices make it possible to assess both the influence of the quality of the emotional expression in the picture and the ability of each participant to recognise emotions. The procedure stems from the assumption that if a picture is not even recognised by the best decoders, then its low hit rate is probably due to the quality of the picture, rather than the decoder.

The results of this study thus provided us with additional criteria for selecting the pictures that would integrate the final set. The criteria to keep the pictures were as follows: (a) the picture should present a high hit rate, preferably above .70 (Ekman & Friesen, 1976), but at least above chance, conservatively set at .33 (e.g., Tracy, Robins, & Schriber, 2009) and (b) the pictures should not present a hit rate higher than chance in non-target emotions. In all cases, the emotion chosen most often corresponded to the target emotion. Furthermore, care was taken that for each emotion, pictures with low and high discrimination indices were retained.

Based on these criteria, 60 pictures were retained in total, 10 (.17) of which were induced expressions and 50 (.83) FACS-posed. The characteristics of the final 60 pictures are presented in Appendix A and examples, in Appendix B. The remaining pictures include expressions by 14 different encoders: 8 males and 6 females. The number of pictures selected for each emotional expression was as follows: 9 for happiness, surprise, disgust, and anger; 8 for sadness; 6 for fear, and 10 for neutral. For an overview of the set broken down per emotion, see Table 3.

Hit rates ranged from .57 to .98 ($M = .86$, $SD = 0.09$; see Appendix A for hit rates per picture). Hit rates were then calculated for emotion by averaging across photographs (Table 4). They ranged from .72 (fear) to .96 (happiness), and are presented in full, along with false alarm rates, in a confusion matrix in Table 5. As can be seen in this table, false alarm

Table 4. Hit rates, Hu scores, and discrimination indices per emotion

	Hit rate			Hu Score			Discrimination index		
Emotion	*M*	SD	95% CI	*M*	SD	95% CI	*M*	SD	95% CI
Happiness	.96	0.04	[.93, .99]	.94	0.05	[.90, 97]	.06	0.04	[.03, .09]
Sadness	.88	0.09	[.81, .96]	.83	0.11	[.73, .92]	.18	0.11	[.08, .27]
Fear	.72	0.09	[.62, .81]	.57	0.05	[.51, .63]	.47	0.16	[.31, .64]
Anger	.88	0.06	[.84, .93]	.81	0.06	[.76, .85]	.19	0.08	[.13, .26]
Surprise	.85	0.11	[.76, .93]	.79	0.12	[.70, .89]	.16	0.10	[.08, .23]
Disgust	.84	0.04	[.81, .87]	.81	0.06	[.77, .85]	.24	0.08	[.18, .30]
Neutral	.88	0.05	[.84, .91]	.82	0.06	[.78, .86]	.22	0.04	[.19, .25]

rates were distributed across non-target emotions in accordance with the similarity between the expressions. For example, pictures of fear were most often confused for disgust and surprise, which is consistent with the literature (e.g., Hawk, van Kleef, Fischer, & van der Schalk, 2009; Langner et al., 2010; Tracy et al., 2009).

Hu scores ranged from .51 to .98 ($M = .81$, SD = 0.11; for Hu scores per picture, see Appendix A and per emotion, Table 4). A 7-level within-subject analysis of variance (ANOVA) was conducted using the Hu scores to compare hit rates per emotion. In order to carry out the analyses, Hu scores were arcsine-transformed. The results revealed a significant main effect of emotion $F(6, 53) = 12.79$, $p < .001$, $\eta^2 = .54$. Tukey-corrected *post hoc* comparisons clearly distinguished three sub-sets. On one hand, the Hu score for fear was significantly lower than that of all the other emotions ($ps < .01$), while happiness was significantly higher ($ps < .05$). The third sub-set, composed of surprise, anger, neutral, disgust, and sadness, was significantly lower than happiness and higher than fear ($ps < .05$).

Discrimination indices ranged from .01 to .66 ($M = .20$, SD = 0.13). They are presented per picture in Appendix A and per emotion in Table 4. Sixty-five per cent of the pictures present moderate-to-high discriminative power, according to Ebel (1965). The FACS scores for each picture can also be found in Appendix A. The degree of correspondence between the AUs of each of the pictures and the target prototypical expression or major variant for that emotion, as presented in Ekman et al. (2002b), was calculated using Wexler's (1972) index of agreement. An average of .92 was obtained for the whole set, ranging from .62 to 1. The FACS-posed pictures included an average of .92 of the prototypical AUs, while the naturally posed pictures included .94.

Discussion

A set of 60 pictures of Argentines expressing the basic emotions was obtained based on their hit rates and discrimination indices. These pictures were all FACS-verified according to the prototypes and major variants established by Ekman et al. (2002b; Ekman, 2007), the FACS scores available for researchers who wish to use the stimuli. Using FACS scores provides emotion researchers with an objective measure to make sure they are dealing with the same facial expressions. Our pictures were not only developed by posing (for the non-induced pictures) the AUs corresponding to the prototypical expressions, but then verified by FACS experts and an agreement index (i.e., Wexler's index). In this sense we can say that these pictures are highly standardised.

The analysis of Hu scores revealed an effect of the specific emotion being expressed on how well that expression was recognised, consistent with previous

Table 5. Confusion matrix: proportion (mean and standard deviation) of perceived emotion per target emotion, averaged across photos

Target emotion \ Perceived emotion	Happiness		Sadness		Fear		Anger		Surprise		Disgust		Neutral		Other[a]	
	M	SD	*M*	SD	*M*	SD	*M*	SD	*M*	SD	*M*	SD	*M*	SD	*M*	SD
Happiness	.96	0.04	.00	0.01	.00	0.00	.00	0.01	.01	0.00	.01	0.02	.01	0.01	.01	0.01
Sadness	.01	0.01	.88	0.09	.01	0.01	.03	0.03	.00	0.00	.00	0.00	.04	0.03	.03	0.02
Fear	.00	0.00	.02	0.01	.72	0.09	.01	0.01	.07	0.05	.01	0.01	.01	0.01	.01	0.02
Anger	.00	0.00	.03	0.05	.02	0.03	.88	0.06	.01	0.01	.08	0.04	.01	0.01	.03	0.02
Surprise	.01	0.00	.01	0.01	.11	0.10	.01	0.01	.85	0.11	.01	0.01	.00	0.00	.03	0.03
Disgust	.00	0.00	.01	0.01	.12	0.09	.02	0.02	.01	0.01	.84	0.04	.00	0.00	.02	0.01
Neutral	.01	0.03	.02	0.03	.01	0.01	.01	0.01	.03	0.05	.03	0.02	.88	0.05	.05	0.04

[a]The category "other" comprised other responses that were incorrect or for which an alternative was not provided, as well as missing responses.

studies. In particular, fear tends to be recognised at lower rates than the rest of the emotions, while happiness tends to be the most easily recognised (e.g., Elfenbein & Ambady, 2002; Ruffman, Henry, Livingstone, & Phillips, 2008), which is also congruent with our findings. Overall, the hit rates and Hu scores observed were similar to those reported in the literature with other stimulus sets (e.g., Table 1; for an overview, see Elfenbein & Ambady, 2002).

Pictures of facial expressions of emotion are commonly used in cognitive and clinical psychology and neuroscience to develop tests that assess the ability to recognise facial expressions of emotions. Considering this tendency, we decided to include in our set pictures that distinguish people with varying levels of ability to recognise emotions. Since there tends to be a negative correlation between hit rates and discrimination indices, a compromise was made in order to retain pictures with high discriminative power, although their hit rates were not the highest.

Study 4. Convergent validity

In order to obtain an additional source of validity, we carried out a study of convergent validity with the POFA, a set with similar characteristics which enjoys widespread acceptance. In a previous study in which the POFA was administered in Argentina (Vaiman et al., 2011), we found hit rates inferior to those reported by Ekman and Friesen (1976) with American participants, which might be explained in part by an in-group advantage. However, in order to demonstrate that our set is more appropriate than the POFA for use locally, we administered both sets to the same sample. We expected to find a high correlation between the two sets, with the UNCEEF outperforming the POFA.

Method

Participants

A total of 212 undergraduate students (71% female and 29% male) from the National University of Cordoba, Argentina, ranging from 16 to 39 years of age ($M = 21$, SD = 4.33) participated in the study.

Measure

The stimuli consisted in the pictures from the UNCEEF and POFA sets. The POFA consists of 110 pictures of Caucasians expressing the basic emotions. The encoders were adults, eight females and six males. For the UNCEEF, the final 60 pictures selected were used. The characteristics of the UNCEEF have been thoroughly described before.

Procedure

The administration of the pictures took place in the same fashion as in Study 3, with each set divided into two sub-sets. The UNCEEF was always presented before the POFA.

Results

Hit rates per picture and emotion were calculated as in Study 3. The data of two participants were excluded from the analyses for presenting recognition rates < = 3.29 SDs overall (Tabachnick & Fidell, 2001) and of five more participants for only having participated in the evaluation of one of the two picture sets.

As in Study 3, hit rates were converted to Hu scores, calculated per emotion for each participant (Schlegel, Grandjean, & Scherer, 2012), which were then arcsine-transformed. Inter-set correlation analyses were carried out between the mean transformed Hu scores per emotion (see Table 6 for the results). Significant correlations between the two sets were found for

Table 6. Results of correlation analyses per emotion between the two sets

Picture set							
	POFA						
UNCEEF	Happiness	Sadness	Fear	Anger	Surprise	Disgust	Neutral
Happiness	.40***	–	–	–	–	–	–
Sadness	–	.30***	–	–	–	–	–
Fear	–	–	.29***	–	–	–	–
Anger	–	–	–	.27***	–	–	–
Surprise	–	–	–	–	.20**	–	–
Disgust	–	–	–	–	–	.31***	–
Neutral	–	–	–	–	–	–	.44***

**$p < .01$.
***$p < .001$.

Table 7. Mean hit rates and Hu scores (and standard deviations) per emotion for each set and results of t tests

Picture set	UNCEEF				POFA						
	Hit rate		Hu score		Hit rate		Hu score				
Emotion	*M*	SD	*M*	SD	*M*	SD	*M*	SD	*t*	*p*	*95%* CI
Happiness	.95	0.04	.92	0.11	.95	0.05	.93	0.09	1.18	.238	[–.02, .07]
Sadness	.86	0.10	.80	0.19	.69	0.20	.65	0.17	10.90	.000	[.23, .33]
Fear	.65	0.22	.53	0.30	.61	0.16	.51	0.23	2.90	.004	[.03, .16]
Anger	.88	0.06	.77	0.18	.82	0.17	.64	0.16	9.55	.000	[.18, .27]
Surprise	.87	0.08	.77	0.16	.94	0.03	.68	0.15	7.21	.000	[.12, .21]
Disgust	.82	0.04	.78	0.19	.77	0.16	.66	0.18	0.21	.000	[.18, .28]
Neutral	.89	0.03	.80	0.18	.85	0.09	.75	0.17	5.38	.000	[.08, .18]

all emotions (*p*s < .001 for all emotions except surprise, where $p < .01$.

A 7 (emotion) × 2 (set) repeated measures ANOVA on recognition accuracy was conducted using the transformed Hu scores, revealing a main effect of set, $F(1, 203) = 150.128$, $p < .001$, partial $\eta^2 = .425$. Descriptive analyses revealed that the pictures of the UNCEEF ($M = .77$, $SD = 0.12$) were recognised to a greater degree than those of the POFA ($M = .69$, $SD = 0.11$). In addition, a main effect of emotion was found, $F(6, 198) = 166.657$, $p < .001$, partial $\eta^2 = .835$. Happiness presented the highest hit rates, followed by neutral, surprise, sadness, disgust, anger, and finally fear. Finally, a set × emotion interaction was observed $F(6, 198) = 17.036$, $p < .001$, partial $\eta^2 = .340$.

In order to explore these effects, paired *t*-tests were conducted between the transformed Hu scores per emotion of the UNCEEF and POFA. Mean hit rates and Hu scores per emotion for each set, as well as the results of the t tests, can be found in Table 7. Significant differences were found between the Hu scores for all emotions (*p*s < .001), except happiness ($p = .24$). For all emotions, the Hu scores of the UNCEEF were higher than those obtained with POFA, except happiness.

Discussion

Both the UNCEEF and the POFA were administered to the same sample of Argentine participants. The results yielded a significant correlation between the two sets, which can be taken as evidence that the two instruments measure the same construct. Moreover, a main effect of set was observed, with the pictures of the UNCEEF being recognised significantly better than those of the POFA per emotion for all emotions except happiness. This could in part be due to an in-group advantage, but may also have to do with the quality of the stimuli (e.g., aesthetics). The results of this study therefore provide convergent validity of the UNCEEF, as well as evidence that it is more suitable than the POFA for evaluating emotion recognition in an Argentine population. Nevertheless, the fact that the UNCEEF was always administered before the POFA constitutes a limitation of this study. Studies involving cognitive tasks usually control the order of presentation of stimuli due to effects of learning or fatigue on performance. However, this was not possible here as the pictures were administered collectively. Therefore, we cannot discard that the order of presentation can have affected the difference in recognition rates between the two sets.

General discussion

Despite a recent tendency to create local sets of pictures of facial expressions of emotion (see Table 1), pictures depicting Latin Americans are still lacking. Here we present the development and validation of a set of pictures of Argentineans expressing the six basic emotions: anger, disgust, fear, happiness, sadness, and surprise, as well as neutral expressions. The set contains 60 pictures of 14 different models of both sexes. The pictures are highly standardised and present contemporary aesthetics. A series of studies was carried out to obtain different sources of evidence for the validity of the set.

First of all, the pictures were FACS-verified by experts, as is becoming standard (e.g., Hawk et al., 2009; Langner et al., 2010; van der Schalk, Hawk, Fischer, & Doosje, 2011; Tracy et al., 2009). An attempt was made to include different prototypical expressions of each emotion and their major variants. The FACS score of each picture is reported, which might be useful for researchers interested in exploring the influence of different AUs on emotion processing. However, while FACS presents the advantage of constituting an objective measure with which to study emotion expression, the ecological validity of posed

expressions of emotion has been questioned (e.g., Russell, 1994). A discussion of this debate exceeds the objectives of the present study, but for a thorough overview, see Russell (1994).

Second, hit rates were obtained for each picture in order to make sure each expression was indeed recognised by non-experts. The hit rates for each picture are provided in order for researchers to be able to choose pictures based on the difficulty desired. The hit rates obtained were relatively high and bore similarity to those of existing sets. Specifically, a mean total hit rate of .87 was obtained, compared to .88 for the POFA (Ekman & Friesen, 1976), .74 for the JACFEE (Matsumoto & Ekman, 1988), .72 for the KDEF (Goeleven et al., 2008), and .82 for the Radboud Faces Database (Langner et al., 2010). However, a couple of pictures were recognised by less than .70 of the participants, which may be explained by the difficulty to express and recognise certain configurations of emotions. As discussed before, the criteria for selecting pictures was not only that they be easily recognised, but also that they be able to discriminate between individuals with more and less ability to recognise emotions. As recognition accuracy and discriminative power tend to be inversely related, this often meant including pictures that did not have the highest hit rates.

Furthermore, our study was the first that we know of to obtain evidence of convergent validity using another set of FACS-based emotion stimuli. Considering the wide acceptance of the POFA, the fact that it significantly correlated with the present set is a good indicator that they measure the same construct. Moreover, the finding that all the emotions expressed in the local set were recognised to a greater degree than those of the POFA, except for happiness, justifies the development of a local set. The differences in performance observed between the two tests may be due to the in-group advantage (e.g., Elfenbein & Ambady, 2002). However, the two stimuli sets differ in more than just the ethnicity of the models. For example, the pictures of the POFA are black and white and were taken several decades ago, so hairstyles and facial hair appear dated. Therefore, in order to test an in-group advantage the same set should be administered to members of different cultural groups (Matsumoto, 2002).

Until now, only two studies that we are aware of have tried to obtain hit rates for emotion pictures in an Argentine sample. One is the study mentioned before with the POFA (Vaiman et al., 2011). The other is a study carried out by Ekman (1972) himself with a set of pictures preceding the POFA, but also of North Americans. The hit rates obtained in those two studies are comparable to those found with the POFA in the present study. Notably, in both Vaiman et al. (2011) and Ekman (1972) fear and sadness were the hardest emotions to recognise, which was also the case with the data obtained with the POFA here. This lends further validity to the findings of this study, and, as all of those studies were carried out with pictures of individuals from the USA, reinforces the importance of a local set of emotion pictures.

Pictures of emotional expressions are often used to create tests to assess the ability to recognise emotions. To this end, we calculated the discriminative index of each photo, and included many pictures with high discriminative power. Discrimination indices may prove useful for researchers interested in emotion recognition ability and are thus reported.

Potential users should however be aware of the limitations of the set. For example, although the set includes many models, we did not manage to include a picture of each emotion for every model and there is not much diversity in the age of the models, with most being young adults. In addition, the set also only contains pictures of the original basic emotions. The pictures were all taken head on, as well, limiting their applicability (see e.g., Langner et al., 2010).

Despite these shortcomings, we hope the set presented here will be of use for future research. What is more, it is our hope that the development of a local set helps promote research on emotion in Latin America. The full set is publicly available and can be accessed from the website https://emocionesunc.wordpress.com or by contacting the authors.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Secretary of Science and Technology of the National University of Córdoba [project code 30820110100062CB].

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational*

and psychological testing. Washington, DC: American Educational Research Association.

Beaupré, M. G., & Hess, U. (2005). Cross-cultural emotion recognition among Canadian ethnic groups. *Journal of Cross-Cultural Psychology*, *36*, 355–370. doi:10.1177/0022022104273656

Biehl, M., Matsumoto, D., Ekman, P., Hearn, V., Heider, K., Kudoh, T., & Ton, V. (1997). Matsumoto and Ekman's Japanese and Caucasian Facial Expressions of Emotion (JACFEE): Reliability data and cross-national differences. *Journal of Nonverbal Behavior*, *21*, 3–21. doi:10.1023/A:1024902500935

Corach, D., Lao, O., Bobillo, C., van Der Gaag, K., Zuniga, S., Vermeulen, M., … Kayser, M. (2010). Inferring continental ancestry of Argentineans from autosomal, Y-chromosomal and mitochondrial DNA. *Annals of Human Genetics*, *74*, 65–76. doi:10.1111/j.1469-1809.2009.00556.x

Ebel, R. L. (1965). *Measuring educational achievement*. Englewood Cliffs, NJ: Prentice Hall.

Ebner, N. C., Riediger, M., & Lindenberger, U. (2010). Faces - a database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior Research Methods*, *42*, 351–362. doi:10.3758/BRM.42.1.351

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

Ekman, P. (2007). The directed facial action task: Emotional responses without appraisal. In J. J. B. Allen (Ed.), *Handbook of emotion elicitation and assessment* (pp. 47–53). New York, NY: Oxford University Press.

Ekman, P., & Friesen, W. V. (1976). *Pictures of facial affect*. Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., & Friesen, W. V. (1978). *The Facial Action Coding System: A technique for the measurement of facial movement*. Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., Friesen, W. V., & Hager, J. C. (Eds.). (2002a). *Facial action coding system: Research nexus*. Salt Lake City, UT: Network Research Information.

Ekman, P., Friesen, W. V., & Hager, J. C. (Eds.). (2002b). *Facial action coding system: Investigator's guide*. Salt Lake City, UT: Network Research Information.

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, *128*, 203–235. doi:10.1037l/0033-2909.1282.203

Frank, M. G., & Stennett, J. (2001). The forced-choice paradigm and the perception of facial expressions of emotion. *Journal of Personality and Social Psychology*, *80*, 75–85. doi:10.1037/0022-3514.80.1.75

Goeleven, E., De Raedt, R., Leyman, L., & Verschuere, B. (2008). The Karolinska Directed Emotional Faces: A validation study. *Cognition & Emotion*, *22*, 1094–1118. doi:10.1080/02699930701626582

Gur, R. C., Sara, R., Hagendoorn, M., Marom, O., Hughett, P., Macy, L., … Gur, E. (2002). A method for obtaining 3-dimensional facial expressions and its standardization for use in neurocognitive studies. *Journal of Neuroscience Methods*, *115*, 137–143. doi:10.1016/S0165-0270(02)00006-7

Hawk, S. T., van Kleef, G. A., Fischer, A. H., & van der Schalk, J. (2009). "Worth a thousand words": absolute and relative decoding of nonlinguistic affect vocalizations. *Emotion*, *9*, 293–305. doi:10.1037/a0015178

Jack, R., Garrod, O., Yu, H., Caldara, R., & Schyns, P. (2012). Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences*, *109*, 7241–7244. doi:10.1073/pnas.1200155109

Jeanneret, S., Rego, P., Oña, A., Vaiman, M., & Pereno, G. L. (2015). Estudio bibliométrico de publicaciones científicas que utilizan pruebas de reconocimiento de emociones faciales. *Anales de Psicología*, *31*, 324–337. doi:10.6018/analesps.31.1.158121

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud faces database. *Cognition and Emotion*, *24*, 1377–1388. doi:10.1080/02699930903485076

Levenson, R. W., Carstensen, L. L., Friesen, W. V., & Ekman, P. (1991). Emotion, physiology, and expression in old age. *Psychology and Aging*, *6*, 28–35. doi:10.1037/0882-7974.6.1.28

Lundqvist, D., Flykt, A., & Öhman, A. (1998). *The Karolinska Directed Emotional Faces*. KDEF, CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institute. ISBN 91-630-7164-9.

Magrassi, G. (1999). *Cultura y Civilización desde Sudamérica*. Buenos Aires: Búsqueda de Ayllu.

Matsumoto, D. (2002). Methodological requirements to test a possible in-group advantage in judging emotions across cultures: comment on Elfenbein and Ambady (2002) and evidence. *Psychological Bulletin*, *128*, 236–242. doi:10.1037//0033-2909.128.2.236

Matsumoto, D., & Ekman, P. (1988). *Japanese and Caucasian Facial Expressions of Emotion (JACFEE) and Neutral Faces (JACNeuF)*. [Slides]. San Francisco, CA: Department of Psychology, San Francisco State University.

Naab, P. J., & Russell, J. A. (2007). Judgments of emotion from spontaneous facial expressions of New Guineans. *Emotion*, *7*, 736–744. doi:10.1037/1528-3542.7.4.736

Rosenberg, E. L., & Ekman, P. (1995). Conceptual and methodological issues in the judgment of facial expressions of emotion. *Motivation and Emotion*, *19*, 111–138. doi:10.1007/BF02250566

Ruffman, T., Henry, J. D., Livingstone, V., & Phillips, L. H. (2008). A meta-analytic review of emotion recognition and aging: Implications for neuropsychological models of aging. *Neuroscience and Biobehavioral Reviews*, *32*, 863–881. doi:10.1016/j.neubiorev.2008.01.001

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin*, *115*, 102–141. doi:10.1037/0033-2909.115.1.102

van der Schalk, J., Hawk, S. T., Fischer, A., & Doosje, B. J. (2011). Moving faces, looking places: The Amsterdam Dynamic Facial Expressions Set (ADFES). *Emotion*, *11*, 907–920. doi:10.1037/a0023853

Schlegel, K., Grandjean, D., & Scherer, K. R. (2012). Emotion recognition: Unidimensional ability or a set of modality-and emotion-specific skills? *Personality and Individual Differences*, *53*(1), 16–21. doi:10.1016/j.paid.2012.01.026

del Sol, M. (2006). Índices faciales en individuos Mapuche. *International Journal of Morphology*, *24*, 587–590. doi:10.4067/S0717-95022006000500012

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston, MA: Allyn & Bacon.

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., … Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, *168*, 242–249. doi:10.1016/j.psychres.2008.05.006

Tracy, J. L., Robins, R. W., & Schriber, R. A. (2009). Development of a FACS-verified set of basic and self-conscious emotion expressions. *Emotion*, *9*, 554–559. doi:10.1037/a0015766

Vaiman, M., Caicedo, E., & Pereno, G. (2011). La expresión de emociones en el Picture of Facial Affect: índices de reconocimiento en una muestra de estudiantes universitarios. *Revista Argentina de Neuropsicología*, *19*, 10–21.

Wagner, H. L. (1993). On measuring performance in category judgment studies on nonverbal behavior. *Journal of Nonverbal Behavior*, *17*, 3–28. doi:10.1007/BF00987006

Wexler, D. (1972). Method for unitizing protocols of descriptions of emotional states. *Journal of Supplemental Abstracts Service. Catalogue of Selected Documents in Psychology*, 2, 116.

Appendix A

Sex of encoder, target emotion, FACS score, hit rate, Hu score, and discrimination index per picture.

Picture	Sex	Emotion	FACS score	Hit rate	Hu score	Discrimination index
C00	Female	Neutral		.89	.83	.23
C01	Female	Fear	1B + 4B + 5D + 16B + 20B + 25B + 38B	.72	.55	.42
C02	Female	Anger	4D + 5B + 23E + 38C	.89	.76	.18
C03	Female	Happiness	6C + 7C + 12C + 25C	.94	.92	.08
C04	Female	Fear	1C + 4C + 5D + 10B + 16B + T23B + 25B + 38B	.83	.66	.27
C05	Female	Surprise	1B + 2B + 5D + 25B + 26B	.79	.68	.23
C06[a]	Female	Fear	1C + R4B + 5E + 10B + 20C + 25C + 26C + 38E	.74	.58	.66
C07[a]	Female	Fear	1B + 4B + 5C + 14B + 20A + 25B	.76	.59	.67
C08	Female	Sadness	1B + 4B + 15B	.69	.58	.30
D00	Female	Neutral		.84	.84	.05
D01	Female	Happiness	6C + 7C + 12D + 25D	.97	.97	.07
D02	Female	Anger	4C + 5C + 17D + 23D + 39C	.97	.84	.18
D05	Female	Disgust	L7B + L9B + 9A + 17C	.85	.84	.18
E00	Female	Neutral		.86	.84	.20
E01[a]	Female	Sadness	1B + 4D + 15B + 17E + 61D	.91	.80	.14
E04	Female	Surprise	1C + 2C + 5C + 25C + 26C	.95	.92	.06
E05[a]	Female	Anger	5B + 7C + 17C + 23D + 38C	.79	.77	.26
F00	Female	Neutral		.86	.84	.26
F02	Female	Sadness	1D + 4D + 15D + 17D + 26B	.90	.87	.23
F03	Female	Surprise	1D + 2D + 5C + 25D + 26D	.95	.94	.07
F04	Female	Disgust	5A + 10C + 15A + 16B + 19 + 25D + 26B	.86	.83	.18
G00	Male	Neutral		.90	.83	.22
G03	Male	Anger	4D + 5D + 23E	.82	.73	.22
G04	Male	Fear	1D + 4C + 5E + 14B + 20A + 25B + 26A	.67	.53	.42
G05[a]	Male	Happiness	6B + 12B	.95	.91	.07
G06	Male	Disgust	10C + 16B + 25C + 26A	.78	.71	.28
H06	Male	Disgust	10B + 15A + 17B + 25C	.89	.88	.22
H08	Male	Surprise	1A + 2B + 5D + 25A	.85	.84	.11
I00	Male	Neutral		.93	.79	.16
I01	Male	Anger	4E + 5C + 7A + 17B + 23C + 38B	.84	.79	.32
I02	Male	Sadness	1C + 4C + 15B + 17C + 38C	.93	.89	.10
I04[a]	Male	Surprise	1A + 2A + 5B + 25C + 26C	.67	.64	.28
I05[a]	Male	Happiness	6C + 7C + 12C + 25D + 26A	.97	.93	.08
K00	Male	Neutral		.77	.75	.24
K01	Male	Sadness	1D + 4B + L4C + 15A + 17D + 38B	.94	.89	.09
K02	Male	Anger	4C + 5B + 24C + 38C	.96	.84	.08
K06	Male	Disgust	4B + 7B + R10C + L10B + 17B + R25B + L25A	.83	.82	.29
L01[a]	Female	Happiness	6C + 7B + 12D + 25C	.98	.98	.05
N02	Female	Happiness	6D + 7D + 12C + 25D	.97	.97	.04
N04	Female	Disgust	10B + 16B + 25C + 26B	.87	.87	.15
O01	Male	Sadness	1D + 4B + 15B	.94	.91	.16
O02	Male	Anger	4D + 7A + 17B + 23E	.91	.76	.22
O06	Male	Fear	1C + 2C + 4B + 5D + 25D + 27D + 53B + 58B + 64B	.58	.51	.40
O08	Male	Disgust	7B + 10B	.84	.78	.28
O10	Male	Happiness	6D + 7D + 12B	.98	.95	.02
O11	Male	Sadness	1A + 4B + 15C + 17B	.82	.79	.32
O12	Male	Surprise	1A + 2A + 5C + 25C + 26C	.93	.68	.13
O13	Male	Disgust	7B + 9B + 17A	.88	.82	.18
P00	Male	Neutral		.89	.86	.18
P01[a]	Male	Anger	4D + R7B + 17D + 24B	.90	.89	.15

(*Continued*)

Appendix A Continued.

Picture	Sex	Emotion	FACS score	Hit rate	Hu score	Discrimination index
P07[a]	Male	Surprise	1C + 2C + 5C	.86	.84	.28
S00	Male	Neutral		.94	.92	.18
S02	Male	Surprise	1D + 2D + 5C + 25D + 27C	.94	.92	.02
S04	Male	Sadness	1E + 4B + 15A + 39B	.95	.89	.03
S05	Male	Happiness	6C + 12C + 25B + 26B	.99	.97	.03
S06	Male	Anger	4D + 5C + 7C + 17C + 24C	.88	.88	.24
U00	Male	Neutral		.90	.70	.20
U03	Male	Happiness	6B + 12B	.87	.82	.15
U04	Male	Surprise	1C + 2C + 5B + 26A	.69	.66	.23
U05	Male	Disgust	10D + 25B + 53B	.76	.74	.40

Note: Picture codes beginning with the same letter were encoded by the same model.
[a]Expressions obtained via emotion induction method.

Appendix B.

Examples of each encoder and emotion.