Photographs of facial expression: Accuracy, response times, and ratings of intensity

ROMINA PALERMO and MAX COLTHEART
Macquarie University, Sydney, New South Wales, Australia

Equal numbers of male and female participants judged which of seven facial expressions (anger, disgust, fear, happiness, neutrality, sadness, and surprise) were displayed by a set of 336 faces, and we measured both accuracy and response times. In addition, the participants rated how well the expression was displayed (i.e., the intensity of the expression). These three measures are reported for each face. Sex of the rater did not interact with any of the three measures. However, analyses revealed that some expressions were recognized more accurately in female than in male faces. The full set of these norms may be downloaded from www.psychonomic.org/archive/.

The information conveyed by facial expressions is so biologically and socially important that humans have evolved complex neural systems that rapidly and accurately decode facial expressions displayed by friends and foes (Rolls, 2000). Over the past few decades, numerous studies have investigated how emotion is recognized from facial expressions (see Adolphs, 2002, for a comprehensive review). Behavioral, neuropsychological, neuroimaging, and electrophysiological techniques have been employed with infants, children, and adults of varying ages. In addition, studies have examined why individuals with specific psychiatric or neurological impairments (e.g., autistic spectrum disorder, Williams syndrome, schizophrenia, Huntington's disease, or Alzheimer's disease) are often impaired in their ability to recognize facial expressions of emotion. In the majority of studies, faces from the Pictures of Facial Affect (Ekman & Friesen, 1976) have been used. This set consists of photographs of five males and six females displaying each of the six basic facial expressions (happiness, sadness, anger, fear, disgust, and surprise) and a neutral expression. Although this database has proved an invaluable resource, use of the database is limited to those researchers who need only a small number of faces.

We collected, from a number of sources, a set of face photographs of 50 individuals displaying basic emotions. Participants were asked to judge the expression displayed by each face and rate the intensity of the expression. Response times (RTs) were also collected. In addition to providing agreement scores and intensity ratings for a larger set of stimuli, we also examined whether there were any differences between male and female expression production and recognition.

Correspondence concerning this article should be addressed to R. Palermo, Macquarie Centre for Cognitive Science (MACCS), Macquarie University, Sydney, New South Wales 2109, Australia (e-mail: rpalermo@maccs.mq.edu.au).

METHOD

Participants

Twelve females ($M = 24$ years, $SD = 8.7$) and 12 males ($M = 25$ years, $SD = 6.6$) participated in the experiment in return for either $10 or course credit.

Stimuli

Source and selection of facial expression stimuli. Faces were sourced from five databases: Pictures of Facial Affect (Ekman & Friesen, 1976), Gur et al. (2002), Mazurski and Bond (1993), NimStim (Tottenham, Borscheid, Ellertsen, Marcus, & Nelson, 2002), and Watson (2001). In this section, we will briefly describe each face set and explain how and why we selected images from each set to be included in this study. To be considered, the photographs needed to be of a Caucasian adult displaying a good representation of *all* seven facial expressions: angry, disgusted, fearful, happy, neutral, sad, and surprised (except for those faces selected from the Gur et al. database, from which surprised expressions were not available).

The Pictures of Facial Affect (Ekman & Friesen, 1976) consist of 110 black-and-white photographs of Caucasian participants. The happy photographs were taken from spontaneous expressions, whereas the participants were asked to move their facial muscles into particular patterns for the other expressions (see Young, Perrett, Calder, Sprengelmeyer, & Ekman, 2002, for further information). Eleven individuals (6 female) were photographed displaying all seven expressions, and these faces were included in the experiment.

Gur et al. (2002) collected three-dimensional color photographs from 139 actors (70 male and 69 female) of diverse ethnicity and age. Half of the expressions were posed, and the other half were evoked by asking the actors to relive appropriate experiences. From this set, we selected a subset of 14 Caucasian individuals (7 female) whose heads were not overly tilted from the vertical and who were between 25 and 51 years of age. Expressions were portrayed in low, medium, and high intensity, and if available, we selected high-intensity expressions for use in this study.

The NimStim Face Stimulus Set (Tottenham et al., 2002) consists of 646 color photographs. Actors of different ethnicities posed each facial expression, and muscles were adjusted until the desired expression was achieved. Each expression was photographed once with the mouth open and once with the mouth closed, except for surprise, for which all the photographs included open mouths. From this database, a subset of photographs of 19 Caucasian individuals (9 female) was selected. Both open- and closed-mouth photographs were used, with selection based on which image appeared to represent the ex-

pression best. A high proportion of open-mouth neutral faces were selected, since no other sets included neutral faces with open mouths.

Mazurski and Bond (1993) collected color photographs from 10 adults (5 female) and 6 children (3 female) who were asked to "show the emotion as they felt appropriate" (p. 41), looking both directly toward and slightly away from the camera. From this set, we selected photographs of 2 adults (1 female) displaying all seven expressions while facing the camera. Black-and-white photographs of 4 individuals (2 female) collected by Watson (2001) were also included.

Facial expression stimuli used in the experiment. To summarize, 25 Caucasian male and 25 Caucasian female individuals were selected. Each individual displayed seven facial expressions: angry, disgusted, fearful, happy, neutral, sad, and surprised (except 7 males and 7 females selected from the Gur et al. [2002] database, where surprised expressions were not available). Thus, a total of 336 faces were included in the experiment (the file AppendixA.xls lists the source and original label for each face).

The faces (256 gray levels, 72 ppi) were scaled so that the interpupillary distance was the same for each face, and they were then cropped to be of equivalent size. If possible, clothing and jewelry were removed using Adobe Photoshop, and any visible background was colored mid-gray. The faces were approximately 5 × 8 cm when displayed on a 17-in. monitor (screen size, 1,024 × 768 pixels). The stimulus presentation was controlled using PsyScript (Bates & D'Oliveiro, 2003) on eMac computers.

Each of the seven expressions was also displayed as a word (*angry, disgusted, fearful, happy, none, sad,* and *surprised*) that the participants were asked to vocalize. Each word was written in 48-point Verdana uppercase in white and was shown on a gray background.

Procedure

Phase 1. The participants were provided with seven emotion terms—angry, disgusted, fearful, happy, none, sad, and surprised—and these options were affixed at the base of the monitor so that they were always visible. The participants were shown one face at a time, in a randomized order on a mid-gray background, and were asked to judge which expression was displayed by the face. The experiment was controlled by a voice-key that collected RTs, and the experimenter recorded the responses. The participants were told to be as accurate as possible and to take as long as they needed to make their responses. A number of practice faces were shown initially to calibrate the voice key threshold. Any faces that were missed due to voice key malfunction were shown to the participant at the end of the session, and their responses (but not RTs) were collected. After judging the expressions, each of the emotion words was presented in the center of the screen five times, in a random order, and the participants were asked to read the word aloud as soon as it appeared.

Phase 2. Seven blocks of trials were formed, with the individual faces in each block being intended to represent the same expression. The participants were told the expression the faces in that block were displaying (e.g., "You will see a number of faces that are *angry*"). They were asked to rate how well the expression was displayed on a 7-point scale (e.g., "Please rate how well you think the expression is displayed using the scale: 1 [*not very angry*] to 7 [*very angry*]"). The order of the seven blocks was organized into six random orders, with 2 females and 2 males receiving each order. Trials in each block were randomized. The rating scale for each block was visible at all times, and the background was mid-gray.

RESULTS

Recognition Rates

The proportion of participants judging each expression to be the one that was intended is shown in AppendixA.xls.

For each participant, mean percent correct scores were calculated for each type of expression and face sex (where each cell mean was based on 18 scores for surprised expressions and 25 scores for each of the other expressions).

A three-way analysis of variance (ANOVA) was conducted with one between-subjects factor (participant sex, female or male) and two repeated measures factors: expression (angry, disgusted, fearful, happy, neutral, sad, or surprised) and face sex (female or male). Significant effects were examined with Tukey HSD tests. There was a main effect of expression [$F(6,132) = 62.54$, $MS_e = 13,316$, $p < .0001$], with happy expressions recognized significantly more accurately ($M = 99.1$, $SE = 0.3$; all $ps < .008$) than all the other expressions. In contrast, fearful expressions were recognized significantly less accurately ($M = 51.7$, $SE = 2.9$; all $ps < .01$) than the other expressions. Neutral ($M = 88.6$, $SE = 1.6$), surprised ($M = 85.0$, $SE = 1.5$), and angry ($M = 82.9$, $SE = 1.2$) expressions were recognized equally well, as were disgusted ($M = 68.2$, $SE = 2.2$) and sad ($M = 62.0$, $SE = 1.8$) expressions. There was also a main effect of face sex [$F(1,22) = 34.31$, $MS_e = 1,267$, $p < .0001$], with expressions recognized more accurately when posed by females ($M = 78.7$, $SE = 1.5$) than when posed by males ($M = 74.8$, $SE = 1.6$).

These effects were moderated by a significant interaction between expression and face sex [$F(6,132) = 7.23$, $MS_e = 385$, $p < .0001$; see Figure 1]. Analyses revealed that both anger (female, $M = 86.7$, $SE = 1.5$; male, $M = 79.2$, $SE = 1.6$; $p < .03$) and sadness (female, $M = 68.8$, $SE = 2.4$; male, $M = 55.2$, $SE = 2.0$; $p < .0001$) were recognized more often in female than in male faces. There were no effects involving participant sex.

We also examined the types of errors made by the participants. The columns in Figure 2 are ordered to indicate perceived similarity between the expressions. The participants were very accurate at recognizing happiness, with less than 1% error. For surprised faces, the participants were most likely to misclassify them as happy (6.5%) or fearful (5.0%). Fearful faces were most often confused with surprised ones (31.1%). Angry faces were most likely to be incorrectly labeled as neutral (5.3%) or disgusted (4.8%). Disgusted faces were most likely to be misclassified as angry (11.8%) or sad (9.9%). For sad faces, the participants were most likely to misclassify them as neutral (17.3%) or disgusted (7.8%). Neutral faces were most often confused with sad ones (4.2%).

Response Times

The average time to vocalize each of the seven emotion terms was calculated for each participant. These times were subtracted from the participant's RTs for correct responses (where the expression was judged to be the same as the intended expression). Incorrect responses were not included in RT analyses. RTs for which a voice key malfunction occurred were omitted (2.4% of the trials). For each participant, mean RTs for each type of expression and face sex were calculated and were based on approx-
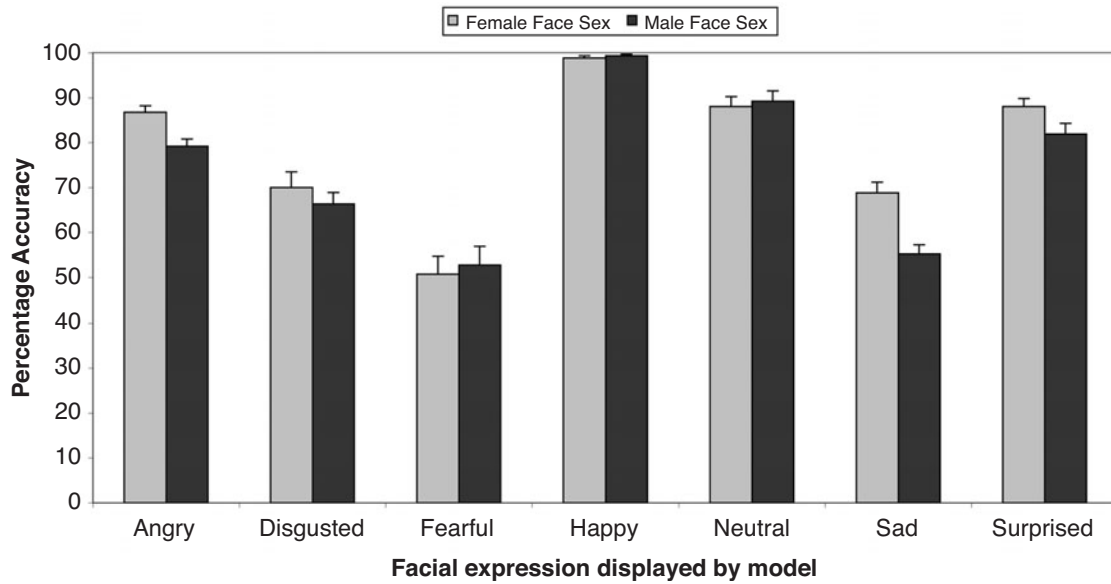
Figure 1. Percentage of expressions displayed by female and male models correctly judged to be the intended expression. Standard error bars are shown.

imately 15 (surprised expressions) and 18 (all other expressions) valid RTs per cell of the design. There were two missing cells, since 1 participant failed to accurately recognize any fearful faces. The mean RTs to correctly recognize each expression are displayed in AppendixA.xls.

As with the accuracy data, a three-way ANOVA was conducted with one between-subjects factor (participant sex, female or male), and two repeated measures factors: expression (angry, disgusted, fearful, happy, neutral, sad, or surprised) and face sex (female or male). There was a main
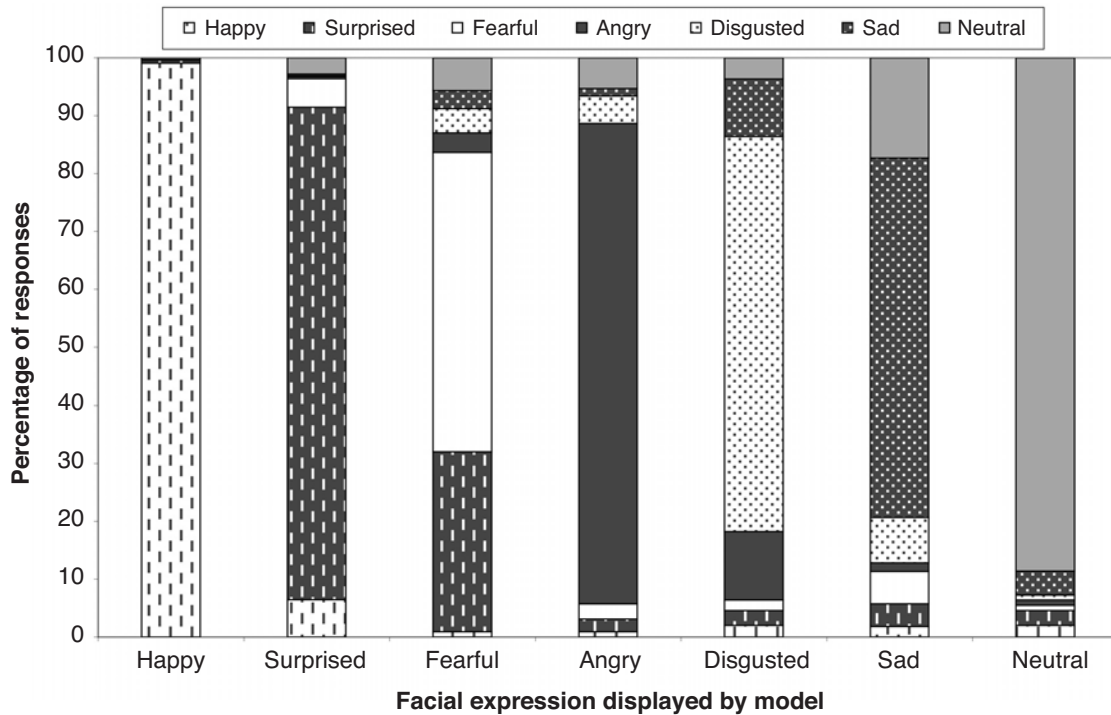
Figure 2. Percentage of responses for each facial expression type, indicating the types of errors made by the participants.

effect of expression [$F(6,132) = 24.4$, $MS_e = 550,476$, $p < .0001$; see Figure 3]. A Tukey HSD test revealed that happy expressions were correctly recognized more quickly than the other expressions ($M = 634$, $SE = 34$; all $ps < .0005$), whereas fearful expressions were recognized significantly slower ($M = 2,389$, $SE = 163$; all $ps < .0002$). There were no effects or interactions involving either face sex or participant sex.

Intensity Ratings

AppendixA.xls lists each face and the corresponding intensity ratings, with higher scores indicating more intense expressions. Pearson correlations were performed for the three dependent variables: intensity rating, accuracy, and mean RT. There was a significant correlation between intensity rating and accuracy ($r = .67$, $p < .0001$), with accuracy increasing as the rating of face intensity increased. There was significant negative correlation between intensity rating and RT ($r = -.54$, $p < .0001$), indicating that as intensity ratings increased, RTs to correctly recognize the intended expression decreased. These patterns of correlation held across all facial expressions except happiness, where the correlations involving rating were not significant (see Table 1). There was also a significant negative correlation between RT and accuracy ($r = -.74$, $p < .0001$), with those faces with higher rater agreement taking less time to identify.

DISCUSSION

For each face, we have collected three measures: the percentage of raters who recognized the intended emotion, the average time taken to recognize the intended emotion, and the average intensity rating. This information (displayed in AppendixA.xls) will be of use to researchers needing to select expressive faces for their studies, because it will allow them to differentiate between faces that clearly display the intended emotion and those that are more ambiguous.

Analyses revealed that happy expressions were recognized both more accurately and more quickly than the other expressions and were rarely confused with other expressions. This finding is consistent with numerous other studies that have found that happiness is the facial expression recognized most rapidly (e.g., Feyereisen, Malet, & Martin, 1986). We also found that fearful expressions were recognized significantly more slowly and less accurately than the other expressions. Russell (1994) also notes that fear is the expression least accurately recognized in forced-choice designs. This may be because fearful faces were often confused with surprised faces, as is often the case in studies in which both fear and surprise labels are used (see Adolphs, 2002, and Rapcsak et al., 2000, for further discussions of this issue).

As was expected, recognizing the intended expression was easiest (as measured by both percentage of rater agreement and RT) when the faces displayed the expression more intensely. Examining the facial expressions separately revealed that this pattern was consistent across all expressions except for happiness, presumably because happiness was recognized at ceiling levels. Similarly, Hess, Blairy, and Kleck (1997) found that recognition accuracy for disgusted, angry, and sad faces increased as the intensity of the expression increased, whereas this was not the case for the recognition of happiness, with even low-intensity happy expressions being accurately recognized.

Overall, expressions were recognized more accurately when displayed by females rather than by males. In particular, both anger and sadness were recognized more often in female than in male faces with this sample of faces. Other studies also have suggested that expressions are more accurately judged when displayed by females rather than by males; however, there does not seem to be any consistent evidence of sex differences for only spe-
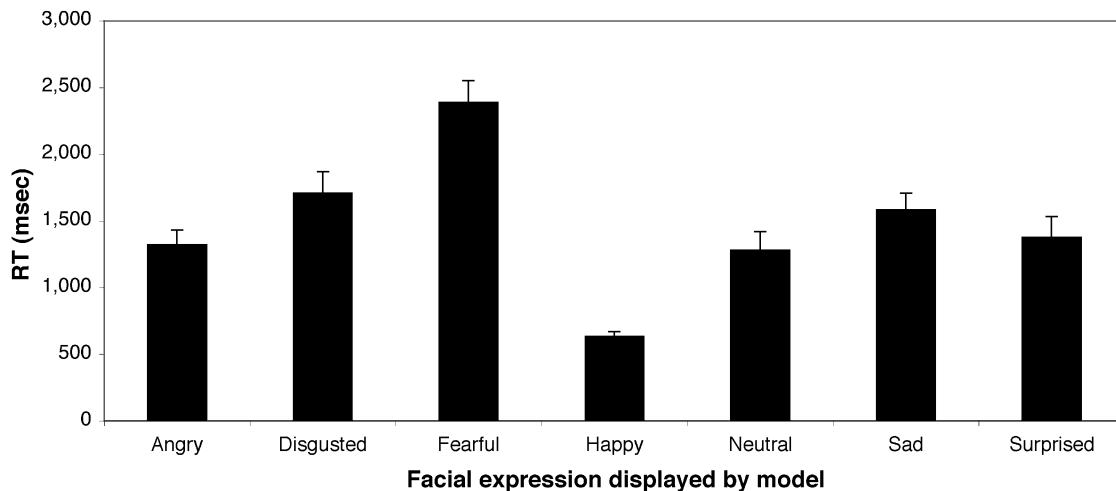

Figure 3. Mean response times (RTs) to correctly recognize each expression. Standard error bars are shown.

Table 1
Pearson Correlations Between Rating, Accuracy, and
Response Time (RT) for Each Facial Expression

Expression	Rating/Accuracy	Rating/RT	Accuracy/RT
Angry	.68***	−.59***	−.64***
Disgusted	.68***	−.62***	−.52**
Fearful	.76***	−.52***	−.42*
Happy	.14	−.21	−.58***
Neutral	.76***	−.59***	−.48**
Sad	.79***	−.56***	−.67***
Surprised	.70***	−.81***	−.72***

*$p < .005$.   **$p < .0005$.   ***$p < .0001$.

cific expressions (Brody & Hall, 1993). Although some studies have shown that females are superior to males in recognizing facial expressions (see reviews by Hall, 1978, 1984), in this study the sex of the rater did not interact with any of the measures. Further studies investigating this issue should use larger sample sizes.

Interest is increasing in understanding the interaction of emotion and cognition. Consequently, it is to be expected that the use of photographs of facial expressions will expand. The measures we report should assist researchers in selecting those faces that are most appropriate to their needs.

OBTAINING THE FACES

A CD containing the Pictures of Facial Affect (Ekman & Friesen, 1976) is available for purchase from http://www.paulekman.com/ for $165. The facial images collected by Gur et al. (2002) are freely available for research purposes and can be requested at http://www.med.upenn.edu/bbl/downloads/index.html. A CD of the Mazurski and Bond (1993) set is freely available via Nigel Bond (n.bond@uws.edu.au). The NimStim face set (Tottenham et al., 2002) is freely available for download from http://www.macbrain.org/faces/index.htm. Development of the MacBrain Face Stimulus Set was overseen by Nim Tottenham and was supported by the John D. and Catherine T. MacArthur Foundation Research Network on Early Experience and Brain Development. Please contact Nim Tottenham at tott0006@tc.umn.edu for more information concerning the stimulus set. The faces collected by Watson (2001) are not available for use by other researchers. Please note that the cost and availability of the face sets were correct at the time of publication.

If researchers have obtained the original photographs from the sources listed above, we will be able to provide the standardized faces used in this study (contact Romina Palermo at rpalermo@maccs.mq.edu.au).

REFERENCES

Adolphs, R. (2002). Recognizing emotion from facial expressions: Psychological and neurological mechanisms. *Behavioral & Cognitive Neuroscience Reviews*, **1**, 21-62.

Bates, T. C., & D'Oliveiro, L. (2003). PsyScript: A Macintosh application for scripting experiments. *Behavior Research Methods, Instruments, & Computers*, **35**, 565-576.

Brody, L. R., & Hall, J. A. (1993). Gender and emotion. In M. Lewis & J. M. Haviland (Eds.), *Handbook of emotions* (pp. 447-460). New York: Guilford.

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

Feyereisen, P., Malet, C., & Martin, Y. (1986). Is the faster processing of expressions of happiness modality specific? In H. D. Ellis, M. A. Jeeves, F. Newcombe, & A. Young (Eds.), *Aspects of face processing* (pp. 349-355). Boston: Nijhoff.

Gur, R. C., Sara, R., Hagendoorn, M., Marom, O., Hughett, P., Macy, L., Turner, T., Bajcsy, R., Posner, A., & Gur, R. 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.

Hall, J. A. (1978). Gender effects in decoding nonverbal cues. *Psychological Bulletin*, **85**, 845-857.

Hall, J. A. (1984). *Nonverbal sex differences: Communication accuracy and expressive style.* Baltimore: Johns Hopkins University Press.

Hess, U., Blairy, S., & Kleck, R. E. (1997). The intensity of emotional facial expressions and decoding accuracy. *Journal of Nonverbal Behavior*, **21**, 241-257.

Mazurski, E. J., & Bond, N. W. (1993). A new series of slides depicting facial expressions of affect: A comparison with the pictures of facial affect series. *Australian Journal of Psychology*, **45**, 41-47.

Rapcsak, S. Z., Galper, S. R., Comer, J. F., Reminger, S. L., Nielsen, L., Kaszniak, A. W., Verfaellie, M., Laguna, J. F., Labiner, D. M., & Cohen, R. A. (2000). Fear recognition deficits after focal brain damage: A cautionary note. *Neurology*, **54**, 575-581.

Rolls, E. T. (2000). Précis of the brain and emotion. *Behavioral & Brain Sciences*, **23**, 177-234.

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.

Tottenham, N., Borscheid, A., Ellertsen, K., Marcus, D. J., & Nelson, C. A. (2002). Categorization of facial expressions in children and adults: Establishing a larger stimulus set [Abstract]. *Journal of Cognitive Neuroscience*, **14**(Suppl.), S74.

Watson, T. (2001). *Perceptual categorisation of anti-expressions.* Unpublished honours thesis, Macquarie University.

Young, A. W., Perrett, D. I., Calder, A. J., Sprengelmeyer, R., & Ekman, P. (2002). *Facial expressions of emotion: Stimuli and tests (FEEST)* [Computer software]. Bury St. Edmunds, U.K.: Thames Valley Test Company.

ARCHIVED MATERIALS