

The similarity of facial expressions in response to emotion-inducing films in reared-apart twins

K. S. Kendler^{1,2*}, L. J. Halberstadt¹, F. Butera¹, J. Myers¹, T. Bouchard³ and P. Ekman⁴

Departments of ¹Psychiatry and ²Human Genetics, Medical College of Virginia, Virginia Commonwealth University, Richmond, VA, USA

³Department of Psychology, University of Minnesota, Minneapolis, MN, USA

⁴Department of Psychiatry, University of California, San Francisco, CA, USA

Background. While the role of genetic factors in self-report measures of emotion has been frequently studied, we know little about the degree to which genetic factors influence emotional facial expressions.

Method. Twenty-eight pairs of monozygotic (MZ) and dizygotic (DZ) twins from the Minnesota Study of Twins Reared Apart were shown three emotion-inducing films and their facial responses recorded. These recordings were blindly scored by trained raters. Ranked correlations between twins were calculated controlling for age and sex.

Results. Twin pairs were significantly correlated for facial expressions of general positive emotions, happiness, surprise and anger, but not for general negative emotions, sadness, or disgust or average emotional intensity. MZ pairs ($n=18$) were more correlated than DZ pairs ($n=10$) for most but not all emotional expressions.

Conclusions. Since these twin pairs had minimal contact with each other prior to testing, these results support significant genetic effects on the facial display of at least some human emotions in response to standardized stimuli. The small sample size resulted in estimated twin correlations with very wide confidence intervals.

Received 17 January 2007; Revised 4 July 2007; Accepted 12 July 2007; First published online 1 October 2007

Key words: Emotion, facial expression, genetics, twin.

Introduction

Self-report measures of many aspects of emotional experience such as happiness, fearfulness, sadness and anger are correlated within families and much of this resemblance appears to result from genetic factors (e.g. Kendler *et al.* 1986, 1994; Goldsmith & Lemery, 2000; Roysamb *et al.* 2003; Rebollo & Boomsma, 2006). A critical methodologic question in the interpretation of these findings is the degree to which self-report measures of emotion are accurate reflections of internal emotional experiences. One way to resolve this question is to develop objective measures of emotional functioning. In this report, we examine one such measure of human emotions: facial expressions.

Darwin (1872) was among the first to recognize the importance of facial expressions in understanding the evolution of emotion. He proposed a relationship between the expressions seen on faces of non-human primates and humans, postulated that the same principles for explaining the morphology of expressions

were applicable to all animals, and suggested that human expressions are universal. On this last topic there has been a great deal of research in the last 40 years, most of which supports Darwin's universality hypothesis (Ekman, 1993; see Ekman, 1999 for a review of the evidence and the challenges to that evidence).

Nested within the universals in the morphology of expressions, however, are large individual differences in the nature and magnitude of facial emotions elicited by the same stimuli (see Ekman, 2007). Although the universality of facial expression of emotion across our species (Ekman, 1993) might suggest a genetic basis for facial displays of emotion, surprisingly little research has examined the genetic basis of normative emotional facial expressions. Prior research has, for example, shown that hypoplasia of the depressor anguli oris muscle (used to turn down the corners of the mouth) is familial (Papadatos *et al.* 1974), that monozygotic (MZ) twins demonstrate greater similarity than dizygotic (DZ) twins in the development of social smiles and intensity and timing of fear reactions in the first year of life (Ekman, 1973) and that adult MZ twins show more similar eye-blink reactions (recorded from the orbicularis oculi muscle) than do DZ twins in response to an acoustic startle when exposed to

* Address for correspondence: K. S. Kendler, M.D., Department of Psychiatry, Medical College of Virginia, Virginia Commonwealth University, Richmond, VA, USA.
(Email: kendler@vcu.edu)

slides of varying emotional content (Carlson *et al.* 1997).

The most relevant prior study examined induced facial expressions in 21 congenitally blind individuals and 30 of their sighted relatives (Peleg *et al.* 2006). Videos were taken and 43 movements analyzed by a single rater. Statistically significant resemblance for specific facial movements was found between the blind individual and the sighted relatives for three emotional states (sadness, anger and 'think-concentrate') but not for three other evaluated emotions (disgust, joy and surprise). The similarities in facial emotions they observed in relatives, the authors argue, are likely due to genetic factors because the blind individuals would not have had the opportunity to learn particular expressions from observing their relatives.

In this report, we examine facial expressions in response to three emotion-inducing films in both members of 28 pairs of MZ and DZ twins from the Minnesota Study of Twins Reared Apart (Bouchard *et al.* 1990). We examine both global measures of emotionality and specific emotional displays obtained from combinations of facial action units coded using the Facial Action Coding System (FACS) (Ekman & Friesen, 1978; Ekman *et al.* 2002). Since these twins were separated early in life and had limited contact with one another prior to testing, twin resemblance should be a result solely of genetic effects. The goal of this study, therefore, was to determine the degree to which genetic factors influence the facial display of emotion in reaction to a standardized emotional stimulus.

Method

Participants

This study, the Emotional Reaction Study, was based on archival data collected from 1979 to 1986 at the Department of Psychology at the University of Minnesota, Minneapolis. The participants were in the Minnesota Study of Twins Reared Apart (Bouchard *et al.* 1990). This large-scale study involved a sample of MZ and DZ twins who were separated in infancy, reared apart, and reunited as adults. Participants completed approximately 50 hours of medical and psychological assessment.

A subset of these twins took part in the Emotional Reaction Study, which was designed by one of us (P.E.), with the goal of assessing the spontaneous expression of emotion. We studied 28 pairs of Caucasian twins (18 MZ, 10 DZ; a set of MZ triplets was counted as three MZ pairs). Zygosity diagnosis was based on extensive serological comparisons, fingerprint ridge count, and anthropometric measurements. There were

16 female twin pairs (9 MZ, 7 DZ) and 12 male twin pairs (9 MZ, 3 DZ). Mean age (s.d.) was 39.0 (9.7) years, with a range of 19–54.

Procedure

Participants were shown three short films chosen by P.E. because of their ability to elicit facial emotional reactions. The first, designed to be pleasant in content, consisted of three 1-minute clips showing, respectively, a gorilla playing in a zoo, ocean waves, and a puppy playing with a flower. The second and third films, designed to be aversive in content, were approximately 2 minutes in duration and consisted, respectively of a safety training film (in black and white) in which two men working carelessly with machines are badly hurt and a medical training film (in color) showing a series of scenes of the treatment of burns and surgery.

The twin participants were given these instructions:

I am going to show you a series of short film clips in order to learn about your reactions. Each film is very short, only a minute or two ... Some of the films bring out positive feelings and some of the films bring out negative feelings. When you are watching one of the negative films you might find yourself feeling quite upset. If the upset feelings ever become too unpleasant for you, remember you can always close your eyes or turn away, or even get up and walk out. Of course we hope you will be able to stay with it, but the decision will be yours.

Prior to the third film, the investigator read an additional instruction to the participant:

The last film is taken from some medical training films. It shows scenes from the treatment of burns and surgery. Some people find this film upsetting. It is just a few minutes. We would like to show it to you. Are you willing to go ahead and look at it?

During exposure to the films, participants' facial expressions were videotaped via a concealed camera the existence of which was known by the participants. Participants sat facing a film screen, with the lower edge of the screen slightly above head level. The camera was positioned so that a head-on, full-face image was obtained while shooting from just below the screen. The camera was approximately 10 feet from the participant and both the camera and the videotape recorder were out of view. Therefore, no observer was visible to the twin during this procedure.

Measurement of facial action

Measurements were made from DVDs that had been transferred from the original black-and-white videotapes. Each of the three films was extracted from the longer videotaped testing sessions and converted to

Table 1. Facial action units used to define facial emotional expressions^a

Emotion	Combination
Positive	(6 and 12) or (12C or 12D or 12E and 6 not present) and 4 not present
Negative	4 or 9 or 10 or 14 or 15 or 20 and 12 not present
Happy	6 and 12 and 4 not present
Surprise	1 and 2 and 5 and 4 not present
Sad ^b	(1 alone) or (1 and 2 and 5 and 15) or (1 and 4) or (1 and 5) or (4 and 15) or (7 and 15) or (11) or (15 and 23 not present) or (17 alone) or (17 and 18 and 23 not present) or (17 and 43 and 23 not present) and 12 not present
Anger	(4 and 5) or (4 and 7) and 12 not present
Disgust	9 or 10 or 16 and 12 not present

^a These numbers refer to the facial action units defined previously by Ekman (Ekman & Friesen, 1978; Ekman *et al.* 2002).

^b All of these codings were proposed *a priori* except for sadness. Our *a priori* codings for sadness occurred too infrequently to usefully analyze and therefore the raters broadened the definition.

mp3 digital format. To keep starting and ending times uniform, each segment extraction began when the subject first appeared to focus on the film being shown and ended with the subject's first facial expression following the film's completion. Using digital video reproduction software, the three segments were assembled with transitions into a single mp3 video and burned onto DVDs. Because no time clock was present on the original videotapes, one was added at this stage to allow coding of onset and offset of facial movements.

All observable facial movement shown by each subject, frame by frame, while watching the films was coded using Ekman and Friesen's Facial Action Coding System (FACS) (Ekman & Friesen, 1978; Ekman *et al.* 2002). The FACS is a comprehensive, anatomically based coding system that describes all visually discernible facial activity in terms of 44 unique action units (AUs), as well as several categories of head and eye positions and movements. Every facial 'event' was scored in terms of the action units that singly or in combination with other action units produced it. Measurement of frequency, duration, and (for some AUs) intensity were scored according to the criteria given in the FACS manual (Ekman & Friesen, 1978; Ekman *et al.* 2002).

Two certified FACS coders (L.H. and F.B.) completed the scoring. Both had completed FACS training and passed the final FACS test, thereby ensuring that their measurements of facial behavior were in good agreement with prior learners of FACS. In each twin pair, one twin's facial activity was initially coded by L.H. and the other twin by F.B. When doing these codings in the 28 pairs, L.H. and F.B. were blind having had no exposure to the film of the co-twin. To obtain inter-rater reliability, both L.H. and F.B. coded the same film on 17 individual twins. For each of the 17 twins, one of the raters was not blind having previously scored the co-twin. These non-blind ratings

were used for the calculation of reliability but not for the examination of twin resemblance.

Because of the small sample size and the rarity of many individual AUs, it was not feasible to analyze meaningfully twin resemblance for every specific AU. Rather, to reduce the number of statistical tests performed and increase the meaningfulness of each test, prior to coding, we developed a system for coding individual AUs into basic emotions based on prior work with the FACS, in particular a table in the FACS manual relating facial actions to emotions, a previously developed table of FACS Affect Codes and the FACS Affect Interpretation Database (FACSAID; Ekman *et al.* 1997). When these various sources disagreed, we used our own judgment to assign AU combinations to emotions beginning first with broad positive *versus* negative emotions. Then, the AUs were broken down into more specific dimensions reflecting six key emotions: happiness, surprise, sadness, fear, anger, and disgust. In the end, our system was similar to but not identical to that developed previously by Ekman (Ekman & Friesen, 1975, 1978; Ekman *et al.* 2002). However, emotional expressions reflecting sadness and fear using our original proposed codings were too rare to be usefully analyzed. Therefore, we expanded these codings. Expressions based on the expanded set of AUs for sadness were now common enough to be analyzed but fear remained too rare to be examined and hence was excluded from these analyses. Table 1 contains the specific AUs and AU combinations that were coded for two general emotion categories and the six specific emotions that were analyzable. (The *a priori* AU codings for sadness were: 1 + 4 + 15 + not 12.)

The onset and offset of each AU was coded which permitted us to examine the display of facial emotions as operationalized by our AU codings in two different ways. First, we examined the number of separate occurrences or count of each emotion. Second, we

examined the total duration of time that the emotion was displayed.

Because the emotional content of film 1 and films 2 and 3 differed substantially, we examined the twin similarity in emotional expression first across all three films and then separately for film 1 (positive emotional content) versus films 2 and 3 (negative emotional content).

In addition, we created a subjective 'global intensity of emotional response' (GIER) rating made for each twin for each of the three films. The rating consisted of a 9-point scale ranging from -4 (extremely negative) to 4 (extremely positive), with 0=neutral (lack of emotional responsiveness). GIER ratings were made by L.H. and F.B. after completing their FACS coding.

Statistical analyses

The distribution of both the count and duration variables for most of the emotions were highly rightward skewed. Pearson product moment correlations were unstable and susceptible to influential data points. We therefore used Spearman rank correlations, controlling for age and sex.

p values are presented one-tailed given the directional hypothesis of resemblance in facial emotional expressions in the twin pairs. While we emphasize results significant at the 0.05 level, we also note, as statistical 'trends', results with *p* values between 0.05 and 0.10. Because intensity was not coded for all facial actions, we only examined intensity averaged over all relevant positive and negative emotional expressions. To reduce the problem of multiple testing, we only examined results separately for MZ and DZ twins if the initial test in all twins had a *p* value ≤ 0.10 .

Reliability

Inter-rater rank correlations ($n=17$) were high for the number of occurrences of the facial expressions (hereafter *count*): general positive +0.81, general negative +0.93 and ranging from +0.81 to 1.00 for specific emotions (all $p < 0.001$). The results were similar for the duration of emotional expressions (general positive +0.85, general negative +0.84 and ranging from +0.85 to 1.00 for specific emotions; all $p < 0.001$). GIER ratings were also highly correlated between the raters (+0.90, $p < 0.001$). Inter-rater correlations for emotional intensity were somewhat lower: general positive +0.59 and general negative +0.51 (both $p < 0.05$).

Results

Global and intensity ratings

The GIER ratings for the three films were significantly correlated in all twins (+0.46, $p=0.01$) and in MZ

(+0.55, $p=0.01$) but not DZ twins (-0.30). The GIER ratings were moderately but non-significantly correlated for their response to film 1 (with positive emotional content) (+0.27) and films 2 and 3 (with negative emotional content) (+0.19). The average intensity of general positive and general negative emotions was not significantly correlated within pairs: -0.33 ($p=0.18$) and +0.03 ($p=0.89$) respectively.

Combinations of facial action coding

Rank correlations and 95% confidence intervals in twin pairs adjusted for age and sex for all films and then separately for film 1 and films 2 and 3 are seen in Table 2 for the count and in Table 3 for duration.

Positive emotions

Examining all films, significant resemblance was seen in twin pairs for the count of general positive emotional expressions and for the specific expressions of happiness and surprise (Table 2). The wide confidence intervals for these and other significant results indicate that, due to the small number of twin pairs examined, the degree of twin resemblance in facial expression of emotion is not known with high precision. Examining film 1 alone, we also saw significant resemblance in the counts of general positive emotions as well as happiness and surprise. No significant twin resemblance was seen for general positive emotions or happiness in response to films 2 and 3. However, a significant twin correlation was seen for expressions reflecting surprise in response to films 2 and 3.

A similar pattern emerged for the duration of the expressions for positive emotions (Table 3) except that the twin correlations were consistently higher than those seen for counts. The durations of general positive emotional expressions were highly correlated in twin pairs in response to all the films and to only film 1 and at a trend level in response to films 2 and 3. A similar pattern was seen for expressions reflecting happiness. As with emotion counts, the duration of facial expressions reflecting surprise was highly correlated in all films and in both film 1 and films 2 and 3 when examined separately.

For counts, correlations were typically higher in DZ than MZ pairs for general positive and happy expressions. By contrast, for expressions of surprise, the MZ twins resembled one another considerably more closely than did the DZ pairs. For duration of emotional expressions, correlations were consistently higher in MZ than in DZ pairs for general and specific positive emotions. Of note, these correlations in MZ pairs consistently exceeded +0.65.

Table 2. Twin similarity for the number of occurrences (counts) of emotional facial expressions^a

Emotion/ Sample	Count		
	All films	Film 1	Films 2 and 3
Positive			
All twins	0.41** (+0.02, +0.68)	0.37** (-0.02, +0.66)	0.25 (-0.15, +0.58)
MZ	0.41* (-0.11, +0.75)	0.25 (-0.28, +0.67)	
DZ	0.56* (-0.23, +0.91)	0.52* (-0.30, +0.89)	
Negative			
All twins	-0.17 (-0.53, +0.23)	-0.12 (-0.49, +0.28)	-0.09 (-0.46, +0.30)
MZ			
DZ			
Happy			
All twins	0.45*** (+0.08, +0.72)	0.40** (+0.01, +0.68)	0.16 (-0.24, +0.52)
MZ	0.33 (-0.20, +0.71)	0.11 (-0.41, +0.58)	
DZ	0.43 (-0.39, +0.87)	0.43 (-0.39, +0.87)	
Surprise			
All twins	0.39** (+0.00, +0.68)	0.46*** (+0.09, +0.72)	0.48*** (+0.12, +0.73)
MZ	0.53** (+0.04, +0.81)	0.68*** (+0.28, +0.88)	0.57*** (+0.11, +0.83)
DZ	0.44 (-0.38, +0.87)	0.10 (-0.65, +0.75)	0.27 (-0.54, +0.82)
Sad^b			
All twins	-0.14 (-0.50, +0.26)	0.00 (-0.39, +0.39)	-0.01 (-0.39, +0.38)
MZ			
DZ			
Anger			
All twins	0.32* (-0.08, +0.63)	0.39** (+0.00, +0.67)	0.35** (-0.04, +0.65)
MZ	0.24 (-0.29, +0.66)	0.14 (-0.38, +0.60)	0.34* (-0.19, +0.71)
DZ	0.13 (-0.63, +0.76)	0.64** (-0.11, +0.93)	0.19 (-0.59, +0.79)
Disgust			
All twins	-0.12 (-0.48, +0.28)	0.03 (-0.36, +0.41)	0.12 (-0.28, +0.48)
MZ			
DZ			

^a Spearman correlations and 95% confidence intervals. All adjusted for sex and age.

^b For sad, unlike the other emotions, *post-hoc* codes were used in analyses because there were too few instances using *a priori* codes.

All one-tailed: * $p \leq 0.10$, ** $p \leq 0.05$, *** $p \leq 0.01$.

Negative emotions

The count for general negative emotions was not significantly correlated in twin pairs. No significant correlations were observed for counts reflecting sadness or disgust to all films or separately to film 1, and films 2 and 3. However, counts of facial expressions reflecting anger were correlated at a trend level in all films and significantly when examined separately in response to film 1, and films 2 and 3.

Examining the duration of facial expressions also showed no twin correlations for general negative emotions or for disgust. At a trend level, the duration of sad expressions was correlated in pairs only in response to film 1. The duration of angry expressions was significantly correlated in pairs again only in response to film 1.

No consistent pattern was seen in the correlations in MZ *versus* DZ twins for counts of angry expressions. MZ twins were more highly correlated than DZ twins in the duration of sad expressions in response to film 1. However, DZ twins were more correlated than MZ twins in the duration of angry expressions in response to the same film.

Discussion

The goal of this study was to determine the role of genetic factors on facial expressions of emotion in response to standardized emotional stimuli. Our study produced four noteworthy findings. First, despite the small sample size, significant twin resemblance was found suggesting that genetic factors do indeed influence emotional facial expressions in a

Table 3. Twin similarity for the duration of emotional facial expressions^a

Emotion/ Sample	Duration		
	All films	Film 1	Films 2 and 3
Positive			
All twins	0.63*** (+0.31, +0.82)	0.60*** (+0.28, +0.80)	0.31* (-0.08, +0.63)
MZ	0.68*** (+0.28, +0.88)	0.75*** (+0.40, +0.91)	-0.07 (-0.55, +0.44)
DZ	0.30 (-0.52, +0.83)	0.30 (-0.52, +0.83)	0.33 (-0.49, +0.84)
Negative			
All twins	-0.34 (-0.64, +0.05)	-0.19 (-0.54, +0.21)	-0.28 (-0.60, +0.12)
MZ			
DZ			
Happy			
All twins	0.63*** (+0.32, +0.82)	0.60*** (+0.28, +0.80)	0.24 (-0.16, +0.58)
MZ	0.73*** (+0.36, +0.90)	0.71*** (+0.34, +0.89)	
DZ	0.37 (-0.45, +0.85)	0.37 (-0.45, +0.85)	
Surprise			
All twins	0.64*** (+0.33, +0.82)	0.46*** (+0.09, +0.72)	0.79*** (+0.58, +0.90)
MZ	0.71*** (+0.33, +0.89)	0.68*** (+0.28, +0.88)	0.84*** (+0.59, +0.94)
DZ	0.56* (-0.24, +0.91)	0.43 (-0.40, +0.87)	0.26 (-0.55, +0.81)
Sad^b			
All twins	-0.10 (-0.47, +0.29)	0.29* (-0.11, +0.61)	-0.05 (-0.43, +0.34)
MZ		0.54** (+0.06, +0.82)	
DZ		-0.02 (-0.72, +0.69)	
Anger			
All twins	0.21 (-0.20, +0.55)	0.36** (-0.03, +0.66)	0.25 (-0.16, +0.58)
MZ		0.30 (-0.23, +0.69)	
DZ		0.63** (-0.13, +0.92)	
Disgust			
All twins	-0.00 (-0.39, +0.38)	0.23 (-0.17, +0.57)	0.10 (-0.30, +0.47)
MZ			
DZ			

^a Spearman correlations and 95% confidence intervals. All adjusted for sex and age.

^b For sad, unlike the other emotions, *post-hoc* codes were used in analyses because there were too few instances using *a priori* codes.

All one-tailed: * $p \leq 0.10$, ** $p \leq 0.05$, *** $p \leq 0.01$.

standardized laboratory situation. Second, the degree of genetic influence on the facial response appeared to vary across emotional categories and was generally more pronounced for positive than for negative emotions. Third, intensity of facial expressions was less reliably rated than the expressions themselves and was not correlated within twin pairs. Fourth, the duration of emotional expressions appeared to be somewhat more influenced by genetic factors than was the number of occurrences of expressions.

Of the 20 correlations for facial expression that were significant at a trend level or greater, follow-up analyses revealed higher correlations in MZ pairs 13 times and DZ pairs seven times. Given that MZ twins share all of their genes while DZ twins on average share only half their genes identical by descent, if

facial emotion were genetically influenced, we would typically expect MZ correlations to exceed DZ correlations. However, this pattern of our results is not unexpected because, given the small sample size, our estimates of twin resemblance are very imprecise. For emotions where the observed DZ correlation exceeded that seen in MZ twins, the confidence intervals were so broad that our findings were still consistent with the hypothesis that the true correlation was higher in MZ pairs. However, because these twins were separated early in life, reared apart and had typically only brief contact with each other prior to testing, even in those situations where the DZ correlations exceed those seen in MZ twins, it is implausible that environmental sources of resemblance would be responsible for the observed patterns.

- facial expression. *Proceedings of the National Academy of Sciences USA* **103**, 15921–15926.
- Rebollo I, Boomsma DI** (2006). Genetic analysis of anger: genetic dominance or competitive sibling interaction. *Behavior Genetics* **36**, 216–228.
- Roysamb E, Tambs K, Reichborn-Kjennerud T, Neale MC, Harris JR** (2003). Happiness and health: environmental and genetic contributions to the relationship between subjective well-being, perceived health, and somatic illness. *Journal of Personality & Social Psychology* **85**, 1136–1146.
- Sullivan PF, Kendler KS, Neale MC** (2003). Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Archives of General Psychiatry* **60**, 1187–1192.
- Sullivan PF, Neale MC, Kendler KS** (2000). Genetic epidemiology of major depression: review and meta-analysis. *American Journal of Psychiatry* **157**, 1552–1562.
- Tellegen A, Lykken DT, Bouchard Jr. TJ, Wilcox KJ, Segal NL, Rich S** (1988). Personality similarity in twins reared apart and together. *Journal of Personality & Social Psychology* **54**, 1031–1039.