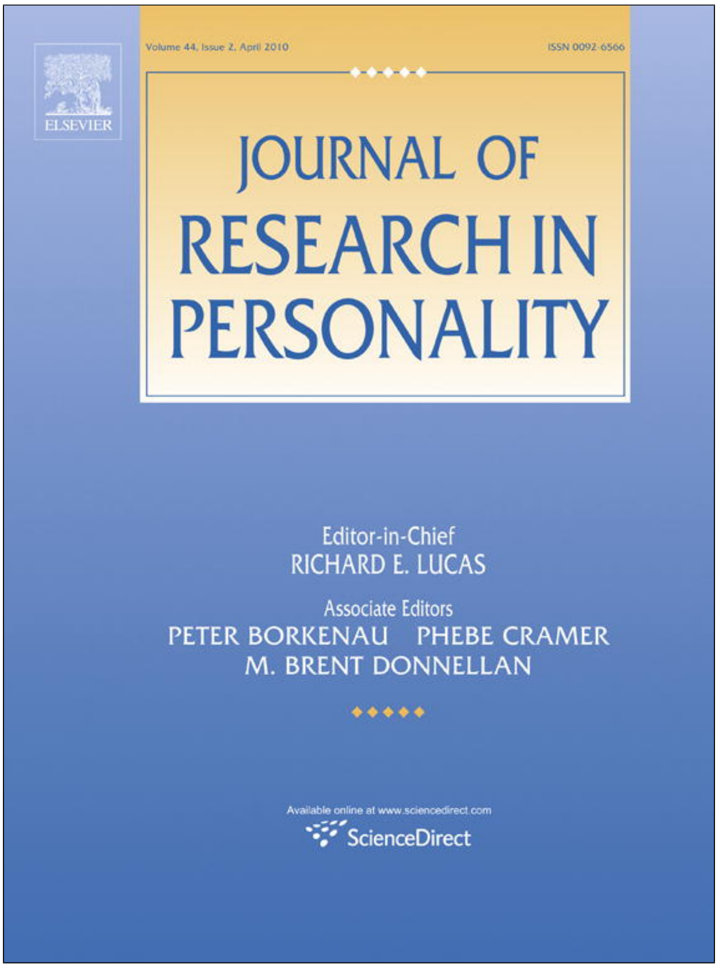
Volume 44, Issue 2, April 2010
ISSN 0092-6566
ELSEVIER
JOURNAL OF
RESEARCH IN
PERSONALITY
Editor-in-Chief
RICHARD E. LUCAS
Associate Editors
PETER BORKENAU PHEBE CRAMER
M. BRENT DONNELLAN
Available online at www.sciencedirect.com
ScienceDirect

Journal of Research in Personality 44 (2010) 199–206

Contents lists available at ScienceDirect

Journal of Research in Personality

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

Individual differences in the accuracy of expressing and perceiving nonverbal cues: New data on an old question

Hillary Anger Elfenbein [a,*], Maw Der Foo [b], Manas Mandal [c], Ramakrishna Biswal [d], Noah Eisenkraft [e], Angeline Lim [f], Sudeep Sharma [a]

[a] Organizational Behavior, Olin School of Business, Washington University in St. Louis, United States
[b] Management and Entrepreneurship Division, Leeds School of Business, University of Colorado at Boulder, United States
[c] Faculty of Social Sciences, Indian Institute of Technology, Kharagpur, India
[d] Department of Psychology, University of Delhi, India
[e] Noah Eisenkraft, Wharton School, University of Pennsylvania, United States
[f] Department of Management and Organization, National University of Singapore, Singapore

ARTICLE INFO

Article history:
Available online 14 January 2010

Keywords:
Emotion
Expression
Perception
Recognition
Social Relations Model

ABSTRACT

Previous research on the link between individual differences in emotional expression and emotion recognition over six decades revealed widely varying results. A recent meta-analysis (Elfenbein & Eisenkraft, 2010) showed a positive correlation for displays elicited as intentional communication, but zero for naturalistic displays. However, the long-standing mystery had dissipated interest, preventing work from using updated authoritative methods for studying individual differences. With Kenny's (1994) Social Relations Model, we tested round robin groups in which each participant posed their emotions and later judged the expressions of each other member. The design included emotion inductions to increase expressers' authentic experience. The resulting effect size, $\rho = .51$, $r = .43$, is larger than previously typical. Implications are discussed for theories on individual emotional skills.

1. Introduction

This study attempts to bring new data to a decades-old mystery, regarding the relationship between individual differences in accuracy of emotional expression and accuracy of emotion recognition. These two skills are central to models of *Emotional Intelligence* (EI; Davies, Stankov, & Roberts, 1998; Mayer, Roberts, & Barsade, 2008; Mayer & Salovey, 1997; Tett, Fox, & Wang, 2005), *Affective Social Competence* (Halberstadt, Denham, & Dunsmore, 2001), *Emotional Competence* (Eisenberg, Cumberland, & Spinrad, 1998; Saarni, 1999), and *Social Skill* (Riggio, 1986). Theoretically, displaying one's emotional cues clearly and perceiving clearly the emotional cues of others are the opposite sides of the same coin.

This question has also received attention across areas of psychology. Those working with conditions such as autism and schizophrenia often use emotional expression and emotion recognition tasks to diagnose and monitor functional deficits (Borod et al., 1990; Halberstadt et al., 2001; Macdonald et al., 1989). Even for individuals of normal or exceptional functioning, emotional skills testing can be worthwhile as a signal of the ability to engage in effective interpersonal relationships, academic achievement, and workplace performance (for reviews, see Elfenbein, Foo, White, Tan, & Aik, 2007; Hall, Andrzejewski, & Yopchick, 2009).

2. New understanding of an old topic

Research on the expression–perception link peaked and then waned many decades ago in the face of conflicting findings—given results ranging all the way from $r = -.80$ (Lanzetta & Kleck, 1970) to $r = +.64$ (Levy, 1964). However, the wide interest in the two underlying constructs ensured that some large-scale data sets over the years still included measures of both, in spite of the loss of empirical attention to their relationship. In attempting to address the long-standing mystery, Elfenbein and Eisenkraft (2010) conducted a meta-analysis and tested several possible moderators of the expression–perception relationship.

One moderator involved the meaning of *accuracy*. That is, some researchers had elicited nonverbal cues as intentional communication displays on the part of participants by asking them to pose—whereas others had coded the clarity of naturalistic behaviors that were elicited while participants were taking part in an emotionally evocative experimental manipulation and unaware they were being observed. This corresponds to a theoretical conceptualization of accuracy that indicates an underlying ability vs. a theoretical conceptualization of accuracy that indicates an underlying

* Corresponding author.
E-mail address: hillary@post.harvard.edu (H.A. Elfenbein).

0092-6566/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jrp.2010.01.001

personality trait, respectively. In keeping with the metaphors of cognitive intelligence and other abilities that are implied by concepts such as Emotional Intelligence and Affective Competence, it would be most consistent to consider the accuracy of posed display as an indication of an individual's underlying ability—that is, displaying poses clearly shows one's skill at using one's nonverbal behaviors deliberately to convey information to others. Carroll (1993) defined ability as "some kind of performance, or potential for performance" (p. 4). He described a metaphor of lifting barbells, and defined ability in this case as the potential to lift a certain amount of weight under favorable conditions—that is, one's maximal performance at a task (notwithstanding that psychometric properties are better when measuring ability in terms of average performance). Carroll (1993) argued that the individual performing an ability task must have some notion of the type of end result and the criterion for being assessed. For this reason, Elfenbein and Eisenkraft (2010) argued that displaying spontaneous emotional expression clearly—in which participants do not have a goal to express their emotions clearly to others—is not a skill, but rather is *legibility*, a personality trait referring to how visibly one reacts as expected to affective stimuli. Thus, it may not be appropriate to consider one's natural level of expressiveness as an ability in conditions for which clarity is not an explicit goal.

This moderating factor was powerful in reconciling past discrepancies. In the meta-analysis, studies of posed emotional expression showed a positive expression–perception link—an average of $r = .19$. However, for those using naturalistic elicitation or a combination of methods, the expression–perception link did not differ reliably from zero.

A second significant moderating factor related to properties of the research design. The meta-analysis showed that most of the remaining heterogeneity among studies vanished when removing those studies that used *round robin* designs—in which each participant in a group judged the affective displays of each other participant. The Social Relations Model (SRM; Kenny & La Voie, 1984) is considered the authoritative statistical technique for analyzing such designs, because it accounts for the interdependence in the data that is inherent in a round robin. After all, each person's display serves as each other's stimulus for perception, and vice versa. However, the studies reviewed in the meta-analyses were conducted before the SRM was developed, so that the researchers had to run conventional analyses by examining the average accuracy without self-ratings. The use of analytic tools that did not account for dyadic interdependence may have introduced some type of bias or error that made these effect sizes among the most variable and extreme in both directions—and suggests that these studies should be placed aside when interpreting the existing body of work (for greater detail, please see Elfenbein & Eisenkraft, 2010).

Taken together, these findings provided a relatively comprehensive answer to the mystery of decades' worth of conflicting empirical findings about the expression–perception relationship. Two significant moderators explained most of the heterogeneity in the pool of studies—in other words, when taking into account one of these factors and when setting aside studies on the basis of the other factor, the apparently inconsistent nature of prior research nearly disappeared.

3. The need for additional research

Meta-analyses are limited by the availability of research for inclusion, and they are also limited by the same limitations of the original underlying empirical work. Notably, the review described above was based on a pool of studies that was somewhat dated, reflecting a spark of interest in the topic during approximately the mid-1960s to mid-1980s, given that researchers tended to lose interest in such an intractable question. This means that research on this topic has not been able to take advantage of recent methodological advances and techniques. In the current paper, we present new data collected in our own laboratories in order to address issues raised by the meta-analysis and to illuminate its findings.

We argue that further research is particularly worthwhile for intentional communication displays—that is, posing—for which the present paper attempts to estimate as precisely as possible the magnitude and confidence intervals of the expression–perception link. This effort has potential theoretical value, because those who have typically grouped together emotional display and perception accuracy into umbrella constructs such as Emotional Intelligence would likely conclude that intentionally communicative display is more relevant to their theoretical models than unobtrusively observed expressivity. After all, Carroll (1993) argued that abilities should be measured in terms of best attempts, rather than the observation of daily life.

3.1. The Social Relations Model

This research uses the Social Relations Model (SRM; Kenny, 1994; Kenny & La Voie, 1984), which was developed for studying any type of individual difference that is generated via inherently dyadic processes—such as interpersonal perception in the communication of emotional cues.

The data generated when examining emotional expression and emotion recognition are, by nature, relational. That is, every measure of an individual's expression ability depends also on the ability of a perceiver to understand this expression. Likewise, every measure of an individual's perception ability depends also on the ability of an expresser to generate an appropriate stimulus for the perceiver to judge. In such an inherently dyadic process, researchers use one of several techniques to analyze individual differences. In the past work reviewed by Elfenbein and Eisenkraft (2010), researchers typically used a separate pool of participants to release themselves from this interdependence. That is, tests of nonverbal perception made use of previously recorded displays for their stimuli, and tests of nonverbal expression made use of outside judges to code the resulting displays. Alternately, researchers used round robin designs—in which each participant perceived the nonverbal cues of each other participant in their group—but, given that the SRM was not available at the time, they analyzed the data using conventional techniques that do not account for dyadic interdependence. As discussed above, these studies yielded findings that were particularly variable.

The final option is to use the SRM, which is considered the best practice for such purposes and includes statistical techniques that account for the interdependence among data points when generated by dyadic processes. The algorithms of the SRM allow researchers to estimate the consistent levels of accuracy that individuals achieve across multiple interaction partners. Consider an example in which Person A makes a judgment of Person B's emotional state, and this judgment is scored for accuracy. The SRM model proposes that this dyadic judgment of emotion is a function of four different components: perceiver effects, target effects, relationship effects, and measurement error. First, the perceiver effect—also known as the actor, decoder, judge, receiver, or recognition effect—is the extent to which A is generally accurate when judging other people's emotions. This is equivalent to an individual difference in emotion recognition skill. Second, the target effect—also known as the expresser, partner, encoder, or sender effect—is the extent to which other people are generally accurate in judging B's emotions. That is, the target effect refers to how clearly B generally expresses himself or herself. Third, the relationship effect is the extent to which A understands B unusually well, after controlling for A's typical perception ability and B's typical expression ability. The relationship term is akin to an interaction term that controls first for main effects. Finally, the fourth component is unstable measurement error. Thus, the algorithms of the SRM take measurements from a dyadic process and provide information on the extent to which they result from two types of systematic individual differences—in this case, emotional expression accuracy and emotion perception accuracy, which are the subjects of the present study.

A further advantage of the SRM is that the tool provides information about the psychometric properties of the emotion recognition and expression measures—and it will not compute associations among variables without significant individual differences to suggest a meaningful level of measurement reliability. In the area of nonverbal communication, it has been notoriously difficult to obtain conventional levels of inter-item reliability. This is because the typical binary scoring of accuracy as correct vs. incorrect gives no partial credit for responses that are partially accurate. Null findings could be common, and effect sizes attenuated, if measures do not reach sufficient reliability for signal to be seen through the noise. For example, the popular Profile of Nonverbal Sensitivity (PONS) test of positive vs. negative and dominant vs. submissive behaviors requires 220 items to reach a Cronbach's alpha of .86 (Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979), which reveals a modest inter-item correlation averaging only $r = .03$ according to the Spearman–Brown equation (Rosenthal & Rosnow, 1991).[1] Given that most past research on the expression–perception link used substantially fewer items, the same inter-item correlation would imply a low effective reliability for measures that could have served to dampen the apparent effect size of the relationship between them. For this reason, we suggest that Elfenbein and Eisenkraft's finding of $r = .19$ is likely to be an underestimate that is attenuated substantially by measurement error (Schmidt & Hunter, 1996), and so we predict a higher value for the present study.

3.2. Other methodological attributes

In addition to using the SRM, we also attempted to specify other design features with the goal of optimizing our ability to estimate the expression–perception relationship for intentionally communicated expressions. First, we tried to increase the natural variation in accuracy levels by including a neutral category in the emotion recognition task (Russell, 1993). Response choices such as 'neutral' or 'none of the above' greatly reduce recognition accuracy away from ceiling effects, because many perceivers endorse correct multiple-choice options they may not believe to fit exactly (Russell, 1993). Notably, having a 'neutral' choice may distract perceivers away from choosing a next-best option when uncertain about the intended emotion in a display,[2] and provides another plausible response for low-intensity displays that are typical except in research that pre-screens stimuli for high intensity. Second, we attempted to increase the variation across individuals by sampling both university students and community members. Third, given sex differences in the accuracy of emotional expression and perception (Hall, 1978, 1984), we recruited a sample balanced by sex.

[1] Note that this analysis of inter-item correlations implicitly presumes that there is one larger-order factor being assessed, which is consistent with researchers' typical use of total scores on the PONS as an indication of individual differences in nonverbal sensitivity (Rosenthal et al., 1979). Indeed, DePaulo and Rosenthal (1979) report that large-scale factor analysis of scores on the PONS reveals one general factor of decoding ability, in addition to four factors related to specific nonverbal channels. Further, they reported that the correlations among various types of decoding skills were "almost always positive" (p. 232). Even so, scores on the PONS test are multi-dimensional, encompassing multiple aspects of nonverbal communication. Thus, inter-item correlations reflect not only measurement error but also meaningful distinctions across heterogeneous types of items.

[2] We thank an anonymous reviewer for this point.

3.3. Emotional experience

In addition to making these methodological changes and updates, we use the present study as a chance to address a theoretical concern that has been raised about the existing pool of existing studies. Cunningham (1977) and others critiqued the display-perception studies that relied on posing because participants had little opportunity to experience affect subjectively and get into a mood with a personal or dramatic context prior to enacting their displays. In these studies with intentionally communicated displays, one could debate whether participants were undergoing any type of affective experience at the time vs. exclusively posing their displays for public consumption. Such posed displays have been critiqued as artificial portrayals that may not capture the richness of authentic emotional expression (Russell, 1994; Russell, Bachorowski, & Fernandez-Dols, 2003). Posed and authentic facial behaviors use separate neuroanatomical pathways, guided by different tracts of facial nerves (see Borod & Koff, 1991; Rinn, 1984). Indeed, clinical case studies show that voluntary and emotional facial movements can each be disturbed by neurological damage that leaves the other intact. For example, brain lesions can limit patients' ability to respond to posing commands without affecting their response to emotion-laden stimuli (Rinn, 1984). Thus, we argue that it is worthwhile to replicate and extend past work in a context that enhances the likelihood that participants had an affective experience at the time their expressions were recorded. In the data presented below, participants took part in a validated emotion induction exercise before posing each state.

4. Overview of the current study

The points raised above are meant to suggest the value of further work on the expression–perception link, which we present below. In keeping with the round robin design of the Social Relations Model (SRM; Kenny, 1994), each participant in our research served as both a perceiver and target, in one of four separate groups of 24 people. Participants took part in two separate sessions: first to record their facial expressions and second to judge the resulting pool of expressions from the other 23 people in their round robin. Given the centrality of *emotion*—within the broader concept of *affect*—to models of emotional intelligence and related constructs, we use established paradigms for measuring emotional expression and emotion recognition. In particular, participants posed and perceived a series of facial expressions. This is because the face is considered the primary canvas for expressing emotion (Ekman & Friesen, 1969), because working with facial expressions is more logistically feasible, and because the past meta-analysis showed no apparent moderation across nonverbal channels of communication.

5. Method

5.1. Participants

Participants were previously unacquainted undergraduate students recruited at a large university in Singapore. In order to maintain a culturally homogeneous sample, we included only those who were born and raised in Singapore, and were of the majority Chinese ethnic group (76.8%; Census Singapore, 2000). The 96 participants included 48 males and 48 females (age $M = 20.3$, $SD = 1.1$, range 19–23). Each was paid S$40, which was the equivalent of approximately US$22, for taking part in both sessions. In Singapore, English is the national language, the language of university instruction, and commonly the language spoken at home; therefore all procedures took place in English.

Due to recruiting that emphasized the multi-part nature of this study, all participants returned for both sessions, resulting in no missing data.

5.2. Display task and procedure

During individual sessions, participants posed displays of six emotional states (anger, disgust, fear, happiness, sadness, and positive surprise), counterbalanced across participants using a Latin square, as well as neutral poses. These are the six emotions identified as basic in Ekman and colleagues' research, on the grounds that they are distinct categories that typically have prototypical signal characteristics that are well recognized (e.g., Ekman, 1972, 1992). We specified surprise to be a positive state in order to limit idiosyncratic interpretation and to include a second emotion of positive valence.

At the end of the emotion induction task described below, participants indicated when they believed they were experiencing the desired affective state. The experimenter told them that the goal was to photograph an example of his or her facial behavior while experiencing this emotion. To limit experimenter bias in choosing a particular moment, participants indicated when they were ready to be photographed by ringing a small bell. As soon as the bell was heard, the experimenter took two photographs of each display, using a digital camera focused on the participant's entire head and neck. Participants then viewed the digital photographs, and had the option to repeat the exercise if they were unsatisfied with their display.[3] To reduce fatigue, participants took a break of at least 3 min after each state. After all six states, the experimenter recorded two neutral expressions, resulting in a total of 14 photographs per participant. The session lasted approximately 1 h.

We elicited examples of affective display with deliberate communicative intent based on Elfenbein, Beaupré, Lévesque, & Hess's (2007) instructions to "Imagine that your close friend is in the room right now, and you want him/her to know how you feel just by your facial expression."

5.3. Emotion induction

Participants took part in an emotion induction for each emotional state they displayed, following Lerner and Keltner's (2001) guided recall procedures. Instructions stated: "Please answer the following two questions as honestly as possible, and also provide as much detail as possible. Please write your description such that someone else reading it might become [emotion] just from learning about the situation. First, describe three to five things that make you most [emotion]. Second, describe in more detail the one situation that makes you, or has made you, most [emotion]."

The guided recall of highly emotional past events reliably predicts self-ratings of subjective emotional experience (Lerner & Keltner, 2001; Levenson, Carstensen, Friesen, & Ekman, 1991; Tsai, Chentsova-Dutton, Freire-Bebeau, & Przymus, 2002), facial behaviors and autonomic nervous system activity corresponding to the emotion recalled (Levenson et al., 1991; Tsai et al., 2002), and effects on factors such as optimism and decision making styles that correspond to dispositional findings for the same emotions (Lerner & Keltner, 2001). Thus, guided recall appears to provide an authentic yet controlled method for eliciting affective states.

A manipulation check was included as independent confirmation of the emotion, to supplement our reliance on past evidence for the validation of similar procedures. After the emotion induction task, participants indicated when they believed they were experiencing the desired affective state. This form of manipulation check was used instead of self-reported emotion scales, because the within-subjects design made it clear to participants that they were intended to feel one particular affective state at a time.

5.4. Judgment task and procedure

In addition to posing their own facial behaviors, as soon as possible after the expression tasks had been completed for all 24 participants in the group, each returned to serve as a perceiver for the resulting collection of displays. Participants viewed the photographs individually using a computerized laboratory task programmed to display each picture one at a time and record a judgment. During the judgment task, each participant viewed all 14 photographs posed by each of the 23 other participants in the same round robin, for a total of 322 test items. Stimuli appeared in a randomized order that differed for each participant. Participants also viewed their own expressions, although the SRM analyses below correct systematically for the influence of missing self-judgments without actually using these data. The judgment task was to select the affective state in the photograph, using a forced-choice of one of seven options: anger, fear, disgust, happiness, neutral, sadness, and surprise. In light of arguments that imposing time limits on ability tests prevents researchers from distinguishing between the rate of work and the overall level of mastery (Carroll, 1993), each photograph remained on the screen until the participant entered a permitted response. To reduce fatigue, participants could take a break at any time, and were encouraged to take a 15-min break at the halfway point. The session lasted approximately 1 h, including breaks.

5.5. Scoring

Because each photograph was generated in the context of a specific intended affective state, the score for judgments were coded dichotomously as 1 for correct and 0 for incorrect, based on whether the response matched the emotion induction after which the display was photographed. Analyses below examine this dichotomous measure of accuracy.

6. Results

Table 1 summarizes the judgment data using a confusion matrix, which plots intended displays across the rows and participant judgments across the columns. The diagonal entries indicate correct judgments. Overall accuracy was 50.8%, whereas guessing due to chance alone would have yielded a score of 14.7%. Scoring these same data as correct vs. incorrect based instead on whether the response matched the general positive vs. negative valence of the intended state, accuracy was 79.4%, which represented a substantial improvement because most confusion took place within the same valence.[4] This suggests that the displays participants produced were recognizable but somewhat subtle. For the purpose of examining individual differences in accuracy, these values are ideal because accuracy shows neither floor nor ceiling effects, even without imposing a time limit.

Although the categorical recognition levels are relatively low compared with other studies of these six emotions (e.g., Ekman, 1972)—which could call into question the expressions produced in this study—we note that it is not meaningful to compare such results directly. First and foremost, Ekman's (1972) stimuli were

[3] This accommodation was intended to increase the comfort level of participants with taking part in the study. Participants almost never accepted the experimenter's offer to repeat the exercise, except in cases where they blinked, sneezed, or were otherwise not fully visible in the photograph.

[4] We thank an anonymous reviewer for pointing out that the general confusion of surprise for the other positive state—happy—rather than for negative states serves as an empirical rationale for calling surprise a positive emotion in this study.

Table 1
communication accuracy and confusion percentages for judgments of facial expressions.

Displayed states	Perceived states							
	Happy (%)	Surprised (%)	Neutral (%)	Afraid (%)	Angry (%)	Disgusted (%)	Sad (%)	Total (%)
Happy	**83.7**	6.3	6.5	0.3	0.4	2.0	0.8	100
Surprised	26.6	**43.0**	14.0	8.2	2.0	4.1	2.0	100
Neutral	4.8	3.7	**59.6**	6.0	7.2	2.7	15.9	100
Afraid	2.0	8.8	15.7	**25.8**	10.9	13.0	23.8	100
Angry	0.8	3.8	18.8	10.0	**35.1**	11.5	20.0	100
Disgusted	3.2	2.2	5.4	13.0	11.4	**50.8**	14.1	100
Sad	0.7	1.2	21.2	6.7	7.8	4.5	**57.9**	100
Category use	17.4	9.9	20.2	10.0	10.7	12.7	19.2	100

Notes: $N = 96$. Bold typeface is used to denote the values in the diagonal cells, which represent the hit rate accuracy coefficients. Values may not add to 100.0% due to rounding. Confusions of 14% or more with an alternate category are underlined.

Table 2
Variance partitioning of accuracy data for the expression and judgment of facial displays of emotion.

Emotion	Expression–perception correlation		Reliability		Source of relative variance				
	Estimate	95% CI	Perceiver	Target	Perceiver (%)	Target (%)	Relationship (%)	Error (%)	Total (%)
Happy	.09	−.11 to .28	.73	.95	3.7*	25.0*	6.6	64.7	100.0
Surprised	.15	−.05 to .34	.57	.96	1.9*	33.4*	5.4	59.3	100.0
Neutral	−.30*	−.46 to −.11	.75	.89	2.2*	12.2*	11.9	70.8	100.0
Afraid	.20*	.00 to .38	.54	.80	8.3*	8.3*	7.0	82.3	100.0
Angry	.22*	.02 to .40	.71	.93	3.1*	23.1*	11.4	61.7	100.0
Disgusted	.20*	.00 to .38	.49	.96	3.2*	33.2*	10.6	54.6	100.0
Sad	−.01	−.21 to .19	.68	.94	3.2*	24.4*	8.8	63.6	100.0
Total	.43*	.27 to .56	.75	.96	4.0*	31.8*	7.9	56.3	100.0

Notes: Reliability refers to the total reliability of individual difference estimates from the SRM model. For variance partitioning, SOREMO provides an indication only of whether or not significance levels reach the cutoff $p < .05$.
* $p < .05$.

preselected to be highly intense and prototypical renditions of the desired emotions. By contrast, the expressions produced in this study may or may not be as intense and/or prototypical. Second, including a neutral category made the judgment test more difficult by providing another plausible response choice for low-intensity displays. Third, the greatest confusions across emotional states tended to occur with other states of the same valence, consistent with past work on emotion perception (e.g., Russell et al., 2003)—which meant that including a second positive emotion also increased the difficulty of the judgment test.

Before analyzing the expression–perception relationship, we first confirmed the psychometric properties of these variables. Table 2 summarizes the results of SRM analyses using the software program SOREMO (Kenny, 1998) to estimate the emotional expression and perception accuracy for each individual participant. Using the 322 test items, the total reliability values were .96 for expression and .75 for perception, which can be interpreted like ordinary Cronbach's alpha coefficients. For perception, these reliability values refer to how much signal is obtained from aggregating the binary accuracy scores across all of the 322 test items that the perceiver judged. That is, the reliability refers to how consistent perceivers were in their performance as good vs. poor judges. Likewise, for expression these reliability values refer to aggregating the total of 322 judgments that the other people in the group made of the participant's expressions (14 photographs judged by 23 people each). That is, the reliability refers to how consistent expressers were in their performance in providing clear vs. unclear portrayals. High values indicate that these measures can be used to describe consistent individual differences in emotional expression and perception ability.

In addition to reporting reliability values that correspond to conventional psychometric statistics, the SRM also provides a *variance partitioning* of the round robin data. This involves determining what proportion of accuracy can be attributed to perceivers, targets, dyads, and measurement error. This variance partitioning can be interpreted akin to an R^2, in that it estimates the variance explained by each of these four mutually exclusive and exhaustive categories. First, perceiver variance refers to the magnitude of individual differences in perception accuracy. Second, target variance refers to the magnitude of individual differences in affective display accuracy. Third, dyadic or 'relationship' effects are akin to an interaction term examining whether some pairs of people may understand each others' expressions better or worse than their individual skill levels would predict. Fourth, error refers to the variance not explained by the three systematic factors, and typically results from the lack of reliability of measures. According to these analyses, all perceiver and target variance estimates reached statistical significance, which suggests that there are substantial enough individual differences in the accuracy of expressing and perceiving emotion to justify testing the relationship between these two processes.

Table 2 also presents estimated coefficients and 95% confidence intervals for the correlation between expression and perception accuracy. Consistent with Elfenbein and Eisenkraft's (2010) meta-analysis, the correlation was significantly positive, $r = .43$, $p < .01$, 95% confidence interval (CI) .27–.56. Disattenuating for measurement error in order to estimate the real effect size in the population (Schmidt & Hunter, 1996) reveals a medium-to-large effect size of $\rho = .51$.[5] As predicted, this effect size is larger than the $r = .20$ typically found in past work that included multiple affective states (Elfenbein & Eisenkraft, 2010). Examining the four round robins as if they were separate studies allowed us to calculate a heterogeneity coefficient of $\chi^2(3) = 1.58$, *ns*, which suggests that

[5] Note that the SOREMO software (Kenny, 1998) provides effects that are already disattenuated by measurement error, and so we multiplied these values by the square root of the product of the two reliability levels when reporting conventional correlations (Schmidt & Hunter, 1996), for comparison with the previous research.

discrepancies in this effect among the four sub-samples was no greater than that expected due to chance sampling.

It is also worthwhile to examine the results for individual affective states. These effects tended to be lower than the effect size for overall accuracy across all seven emotional states tested—most likely due to the lower reliability of the individual affective state ratings. That is, composites have greater reliability than the individual subscales of which they are composed. For the most part, the correlations had positive trends that reached statistical significance for anger, disgust, and fear. For neutral expressions, there was a significant negative correlation, $r = -.24$, $p = .02$. We speculate that this could result from the possible usage of neutral response to indicate 'none of the above' or 'uncertain', in which case apparent neutral perception accuracy could result from the frequent endorsement of the neutral category—rather than necessarily from better performance.

7. Discussion

This paper presents the first new empirical research in decades on a question that has had longstanding importance in the study of individual differences: the association between accuracy in emotional expression and emotion recognition. This topic was an unsolved mystery in the field of nonverbal behavior and person perception—after capturing extensive attention for several decades that ended about 25 years ago—and appears to have been largely resolved by a recent meta-analysis (Elfenbein & Eisenkraft, 2010). Even so, meta-analysis is limited by the availability of past research, and it is limited by the same limitations as the underlying studies. Notably, the loss of interest over time in such an apparently intractable question meant that the pool of studies was too dated to incorporate new advances in methodological best practices. Thus, the present work was intended to confirm and extend these past results.

Using Kenny's Social Relations Model (SRM; Kenny, 1994), we replicated and illuminated the findings of the meta-analysis. Notably, we found a medium-to-large positive correlation between emotion recognition accuracy and the accuracy of deliberately posing emotional expressions. The coefficient of $r = .43$ ($\rho = .51$) is substantially greater than the average $r = .19$ in past work. We predicted that the present study would yield a greater effect size than previous studies due to a combination of factors, including the precision of the SRM methods and analytical techniques, a large number of repeated measures, stimuli that were likely more intense due to the use of emotion induction exercises during the expression protocol (Cunningham, 1977), and increased variation in accuracy levels by including a neutral category in the emotion perception task (Russell, 1993). This extra precision is important because the magnitude of association between these two individual difference factors is important to clinicians and theorists working in the area of emotional skills.

7.1. Limitations and future research

Although we argue that the additional work from our laboratory adds to the existing body of evidence, this study itself is not without limitations. Notably, this research examined only still facial photographs, and additional research should include emotional cues expressed via nonverbal channels such as facial movement, vocal tone, and body language. In addition, the forced-choice design seemed necessary due to the two-part protocol—in which participants were already aware of which affective states had generated the stimulus materials when they took part in judgment sessions—but the merits of such designs requiring verbal response have been debated (e.g., Barrett, Lindquist, & Gendron, 2007; Frank & Stennett, 2001; Russell, 1993). There are tradeoffs involved in all types of participant response, in that open-ended responses require extensive coding and researcher judgment, and dimensional responses are highly time-consuming for participants. Thus, it is valuable use a range of response types where possible, rather than to depend on any one type to the exclusion of the others. Further, we specified surprise to be a positive state in order to limit idiosyncratic interpretation and to include a second emotion of positive valence, but prototypical accounts consider surprise to have a neutral valence or a valence determined by personal preferences regarding the experience of surprise (Ekman, 2004). Another limitation is that participants may have experienced fatigue during the lengthy protocols, which could have influenced their expression or recognition accuracy. Further, our method of administering an affect induction exercise to participants before they posed facial expressions means that the study cannot be interpreted as having produced fully posed vs. fully spontaneous expressive displays.

As with most research on individual differences, we sampled participants from only one nation. Given accumulating evidence for cultural differences in the expression and perception of emotions (for reviews, see Elfenbein & Ambady, 2003; Mesquita, Frijda, & Scherer, 1997), it is worthwhile to conduct research on these topics in a variety of cultural settings. Although it is possible that the relationship between emotional expression and recognition accuracy may vary in other cultures, we note that the results reported here from Singapore are consistent with those from previous work conducted in Western cultures.

Another limitation of the present research is that it sheds light on the relationship between emotional expression and perception only for posed emotions. Although some scholars have gone so far as to question the validity of all research using freely posed facial expressions (c.f., Matsumoto, Olide, & Willingham, 2009), we note that intentional posing is a common practice within research on affective cues (Juslin & Scherer, 2005). In the present research, we used posing not for the sake of methodological convenience, but rather because posing corresponds most closely to theorists' operational definitions of ability (Carroll, 1993). Even so, we also conducted a pilot test to examine whether use of updated research methods would yield similar results to the meta-analysis when examining an expressive condition that was more naturalistic—for which Elfenbein and Eisenkraft (2010) had found a cumulative effect size of zero. We collected additional data using culturally homogeneous samples of 84 people from each of the Punjab and Orissa regions of India, who took part in round robins of 12 people each, for a total $N = 168$. These participants followed the same protocol described above, with the exception that their facial expressions were collected in the context of a pseudo-naturalistic rather than deliberately expressive protocol. Because a truly spontaneous condition was not feasible in this setting, we adapted a naturalistic protocol from Bonanno, Papa, Lalande, Westphal, and Coifman (2004), which informed participants that they were sitting in front of a recording device, but in one condition were instructed "Simply view the images in any way you would naturally do so." Accordingly, instead of telling our participants to pose so that an observer would know how they feel from their facial expression alone—as in the data presented above—these participants were asked to act naturally. This protocol was pseudo-naturalistic in the sense that participants were asked to behave normally rather than to attempt to be clearly expressive, but they were aware of the presence of an observer. In these data, consistent with the recent meta-analysis, our estimate of the expression–perception relationship was $r = .03$, *ns*, 95% CI $-.13$ to $.18$, which could not be distinguished from zero. This suggests that it is not the use of the Social Relations Model or other updated methods alone that provided such a large effect size in the present study, but rather the use of these methods in the context of intentionally communicated expressions.

Ultimately, we examined the relationship between only two emotional skills among the many that have been linked together within umbrella models of Emotional Intelligence (e.g., Davies et al., 1998; Mayer, Salovey, & Caruso, 2000). We began with emotional expression and emotional perception because they are the flip sides of the same coin—the clear communication of affective states via nonverbal cues—and because they had a decades-old history of research available for review. However, to provide further grounding for theories about individual differences in Emotional Intelligence, future work should examine the convergence among additional emotional skills, such as emotion regulation, intrapersonal awareness, and managing the emotions of others. In the meantime, the present findings offer encouraging evidence that the operational definition of emotional expression accuracy most consistent with ability models (Carroll, 1993)—i.e., the clarity of deliberate expressions—correlates significantly with emotion recognition. This suggests the intriguing possibility that distinct emotional skills may converge with each other.

7.2. Practical implications

In addition to the theoretical implications discussed above, these findings have important practical implications for clinicians and researchers. Individual differences in communicating emotional states via nonverbal cues are associated with important personal outcomes as broad as social adjustment, mental health, academic achievement, and workplace performance (Elfenbein et al., 2007; Hall et al., 2009; Nowicki & Duke, 1994; Rosenthal et al., 1979)—which means that the phenomenon also touches clinical, cognitive, developmental, and industrial/organizational psychology. The measurement of emotional expression and emotion recognition can assist healthcare professionals in monitoring the prognosis and treatment of conditions such as schizophrenia and autism, school counselors in monitoring the social adjustment of children, and, more recently, industrial psychologists in measuring the emotional skills of current and prospective employees for selection, training, and evaluation.

Thus, it is worthwhile to understand how much unique information such professionals obtain from employing diagnostics of both emotional display and perception. Tests of emotion recognition are relatively easy and inexpensive to administer, given that standardized scales have been developed in which samples of nonverbal behavior are displayed to participants, who typically enter multiple-choice judgments that can be scored using a simple answer key. For example, the popular DANVA perception test (Nowicki et al., 1994) takes participants approximately 5 min each for facial expression, vocal tone, and body posture subscales. By contrast, tests of emotional expression are relatively challenging and expensive to administer. Participants need to take part in an activity that is designed to elicit and record samples of their expressions. Testers then need to code the resulting expressions for their clarity by a panel of outside judges, or using specialized systems such as the Facial Action Coding System (FACS; Ekman & Friesen, 1978) that require extensive training. This difference in effort makes it clear that emotional expression accuracy is only worth measuring when it provides sufficiently non-redundant information with respect to emotion recognition. The current findings suggest that, when the stakes are sufficiently low, such efforts are not worthwhile for those interested in the clarity of purposeful display—for which the association with perception accuracy appears to be consistently positive and reasonably large. By contrast, such efforts seem to be worthwhile for those who are interested in everyday functioning rather than capability and, thus, the clarity of naturalistic displays—for which perception accuracy provides no apparent information. In such cases, researchers or clinicians might also consider unobtrusive observation or assessments from peers or family members of the participants or patients—as an inexpensive alternative that can be validated as an indicator of the legibility of naturalistic emotional expression. Thus, depending on a clinician's goals, these two complementary processes may or may not both be worth measuring.

Acknowledgments

The second and third authors contributed equally and appear in alphabetical order. The fourth, fifth, six, and seventh authors contributed equally and appear in alphabetical order.

Preparation of this article was supported by National Science Foundation Grant BCS-0617634 to Hillary Anger Elfenbein and Faculty Research Grant RP-317-000-053-112 from the National University of Singapore to Maw Der Foo.

We thank Jennifer Boldry, Stuart Bunderson, Stéphane Côté, Kurt Dirks, Ursula Hess, Gavin Kilduff, David Matsumoto, Jeffrey Sanchez-Burks, and Gerard Saucier for helpful comments, and Rahul Chaturvedi, Kang-Wei Chong, Somnath Dutta, Kripa Shankar Gupta, Victor Lau, Abha Singh, and Wee-Leng Thye for research assistance.

References

Barrett, L. F., Lindquist, K., & Gendron, M. (2007). Language as a context for emotion perception. *Trends in Cognitive Sciences, 11*, 327–332.

Bonanno, G. A., Papa, A., Lalande, K., Westphal, M., & Coifman, K. (2004). The importance of being flexible: The ability to enhance and suppress emotional expression predicts long-term adjustment. *Psychological Science, 157*, 482–487.

Borod, J. C., & Koff, E. (1991). Facial asymmetry during posed and spontaneous expression in stroke patients with unilateral lesions. *Pharmacopsychoecologia, 4*, 15–21.

Borod, J. C., Welkowitz, J., Alpert, M., Brozgold, A. Z., Martin, C., Peselow, E., et al. (1990). Parameters of emotional processing in neuropsychiatric disorders—Conceptual issues and a battery of tests. *Journal of Communicative Disorders, 23*, 247–271.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.

Cunningham, M. R. (1977). Personality and the structure of the nonverbal communication of emotion. *Journal of Personality, 45*, 564–584.

Davies, M., Stankov, L., & Roberts, R. D. (1998). Emotional intelligence: In search of an elusive construct. *Journal of Personality and Social Psychology, 75*, 989–1015.

DePaulo, B. M., & Rosenthal, R. (1979). Ambivalence, discrepancy, and deception in nonverbal communication. In R. Rosenthal (Ed.), *Skill in nonverbal communication* (pp. 204–248). Cambridge: Oelgeschlager, Gunn & Hain.

Eisenberg, N., Cumberland, A., & Spinrad, T. L. (1998). Parental socialization of emotion. *Psychology Inquiry, 9*, 241–273.

Ekman, P. (1972). Universals and cultural differences in facial expressions of emotion. In J. Cole (Ed.). *Nebraska symposium on motivation, 1971* (Vol. 19, pp. 207–282). Lincoln: Lincoln.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6*, 169–200.

Ekman, P. (2004). *Emotions revealed: Recognizing faces and feelings to improve communication*. New York: Holt.

Ekman, P., & Friesen, W. V. (1969). Nonverbal leakage an clues to deception. *Psychiatry, 32*, 88–106.

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

Elfenbein, H. A., & Ambady, N. (2003). Universals and cultural differences in recognizing emotions. *Current Directions in Psychological Science, 12*, 159–164.

Elfenbein, H. A., & Eisenkraft, N. (2010). The relationship between displaying and perceiving nonverbal cues of affect: A meta-analysis to solve an old mystery. *Journal of Personality and Social Psychology, 98*, 301–318.

Elfenbein, H. A., Beaupré, M. G., Lévesque, M., & Hess, U. (2007). Toward a dialect theory: Cultural differences in the expression and recognition of posed facial expressions. *Emotion, 7*, 131–146.

Elfenbein, H. A., Foo, M. D., White, J. B., Tan, H. H., & Aik, V. C. (2007). Reading your counterpart: The benefit of emotion recognition ability for effectiveness in negotiation. *Journal of Nonverbal Behavior, 31*, 205–223.

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.

Halberstadt, A. G., Denham, S. A., & Dunsmore, J. C. (2001). Affective social competence. *Social Development, 10*, 79–119.

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

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

Hall, J. A., Andrzejewski, S. A., & Yopchick, J. E. (2009). Psychosocial correlates of interpersonal sensitivity: A meta-analysis. *Journal of Nonverbal Behavior, 33*, 149–180.

Juslin, P. N., & Scherer, K. R. (2005). Vocal expression of affect. In J. A. Harrigan, R. Rosenthal, & K. R. Scherer (Eds.), *The new handbook of methods in nonverbal behavior research*. Oxford University Press.

Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis*. New York: Guilford Press.

Kenny, D. A. (1998). SOREMO [computer program]. <http://nw3.nai.net/~dakenny/kenny.htm>.

Kenny, D. A., & La Voie, L. (1984). The social relations model. *Advances in Experimental Social Psychology, 18*, 141–182.

Lanzetta, J. F., & Kleck, R. E. (1970). The encoding and decoding of nonverbal affects in humans. *Journal of Personality and Social Psychology, 16*, 12–19.

Lerner, J., & Keltner, D. (2001). Fear, anger, and risk. *Journal of Personality and Social Psychology, 81*, 146–159.

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.

Levy, P. K. (1964). The ability to express and perceive vocal communication of feeling. In J. R. Davitz (Ed.), *The communication of emotional meaning*. New York: McGraw-Hill.

Macdonald, H., Rutter, M., Howlin, P., Rios, P., Le Conteur, A., Evered, C., et al. (1989). Recognition and expression of emotional cues by autistic and normal adults. *Journal of Child Psychiatry and Psychology, 30*, 865–877.

Matsumoto, D., Olide, A., & Willingham, B. (2009). Is there an ingroup advantage in recognizing spontaneously expressed emotions? *Journal of Nonverbal Behavior, 33*, 181–192.

Mayer, J. D., Roberts, R. D., & Barsade, S. G. (2008). Human abilities: Emotional intelligence. *Annual Review of Psychology, 58*, 507–536.

Mayer, J. D., & Salovey, P. (1997). What is emotional intelligence? In P. Salovey & D. J. Sluyter (Eds.), *Emotional development and emotional intelligence: Educational implications* (pp. 3–25). New York: Basic Books.

Mayer, J. D., Salovey, P., & Caruso, D. R. (2000). Models of emotional intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 396–420). Cambridge, England: Cambridge University Press.

Mesquita, B., Frijda, N. H., & Scherer, K. R. (1997). Culture and emotion. In J. W. Berry, P. R. Dasen, & T. S. Saraswathi (Eds.). *Handbook of cross-cultural psychology* (Vol. 2, pp. 255–297). Boston: Allyn and Bacon.

Nowicki, S., Jr., & Duke, M. P. (1994). Individual differences in the nonverbal communication of affect: The diagnostic analysis of nonverbal accuracy scale. *Journal of Nonverbal Behavior, 19*, 9–35.

Riggio, R. E. (1986). Assessment of basic social skills. *Journal of Personality and Social Psychology, 51*, 649–660.

Rinn, W. E. (1984). The neuropsychology of facial expression: A review of the neurological and psychological mechanisms for producing facial expressions. *Psychological Bulletin, 95*, 52–77.

Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., & Archer, D. (1979). *Sensitivity to nonverbal communication: The PONS test*. Baltimore: Johns Hopkins University Press.

Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd ed.). New York: McGraw-Hill.

Russell, J. A. (1993). Forced-choice response format in the study of facial expression. *Motivation and Emotion, 17*, 41–51.

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.

Russell, J. A., Bachorowski, J., & Fernandez-Dols, J. (2003). Facial and vocal expressions of emotions. *Annual Review of Psychology, 54*, 329–349.

Saarni, C. (1999). *The development of emotional competence*. New York: Guilford Press.

Schmidt, F. L., & Hunter, J. E. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods, 1*, 199–223.

Tett, R. P., Fox, K. E., & Wang, A. (2005). Development and validation of a self-report measure of emotional intelligence as a multidimensional trait domain. *Personality and Social Psychology Bulletin, 31*, 859–888.

Tsai, J. L., Chentsova-Dutton, Y., Freire-Bebeau, L., & Przymus, D. E. (2002). Emotional expression and physiology in European Americans and Hmong Americans. *Emotion, 2*, 380–397.