Facial Emotion Recognition Accuracy and Child Physical Abuse: An Experiment and a Meta-Analysis

Michael F. Wagner, Joel S. Milner, Randy J. McCarthy, Julie L. Crouch, Thomas R. McCanne, and John J. Skowronski

Center for the Study of Family Violence and Sexual Assault, Northern Illinois University

Objective: To examine child facial emotion recognition accuracy (ERA) in high-risk for child physical abuse (CPA) parents and low-risk for CPA parents (Study 1) and to conduct a meta-analysis summarizing published research on the relationship between child facial ERA and CPA (Study 2). Method: In Study 1, ERA data for child facial emotions were obtained from mothers and fathers who were at high risk (n =51) or low risk (n = 61) for CPA. In 1 of 2 presentation time conditions (100 ms, 600 ms), parents evaluated child photographs, taken from the Radboud Faces Database, which displayed 5 face emotions (angry, happy, sad, fearful, neutral) at 3 face angles (frontal, 45 degrees, 90 degrees). In Study 2, a meta-analysis of published studies was used to estimate the overall effect size of ERA differences between high-risk/abusive and comparison parents. Results: In Study 1, ERA differences were found for emotions (largest ERA for happiness), face angles (frontal > 45 degrees > 90 degrees), and presentation times (100 ms < 600 ms); however, only an overall trend for ERA risk group differences was observed. Nevertheless, the Study 2 meta-analyses revealed a significant effect size reflecting an overall moderate ERA difference between high-risk/abusive and comparison parents, and the effect size was not moderated by the population studied (high-risk parents vs. abusive parents). Conclusions: Because child facial ERA appears to be associated with CPA, the manner in which parental child emotion recognition errors contribute to problematic parent-child interactions merits additional study.

Keywords: child physical abuse, child abuse risk, facial emotion recognition, facial emotion recognition accuracy, meta-analysis

The competent regulation of one's emotions is believed to play a major role in the development and maintenance of social competence, including the ability to engage in effective parenting behaviors (Izard, 1971); whereas emotion regulation deficits are thought to be associated with interpersonal conflict and aggression, including problems in parenting behaviors (Dix, 1991). Although emotion regulation requires multiple skills (Dix, 1991; Shields & Cicchetti, 1998), models of emotional competence (e.g., Saarni, 1999) and models of emotional intelligence (e.g., Mayer, Roberts, & Barsade, 2008) suggest a necessary first step in the development of emotion regulation is learning to recognize emotions in others. The present research (Study 1) was designed to evaluate the extent to which high-risk for child physical abuse (CPA), compared to low-risk for CPA, parents have emotion recognition deficits when viewing chil-

published studies, the present research (Study 2) sought to estimate the overall weighted mean effect size (ES) for child facial emotion recognition accuracy (ERA) differences between abusive/high-risk parents and nonabusive/low-risk parents and to determine if child facial ERA varies as a function of the population studied (abusive parents vs. high-risk parents).

Research indicates that the relationship between emotion recog-

dren's facial emotions. In addition, based on a meta-analysis of

Research indicates that the relationship between emotion recognition, social competence, and aggression is complex. Studies have linked ERA to the development of social competence in children and adolescents; whereas certain types of emotion recognition differences (e.g., a negativity bias defined as a tendency to see neutral and ambiguous facial stimuli as representing negative emotions, such as anger), as opposed to overall ERA deficits, have been associated with behavior problems including aggression (see Trentacosta & Fine, 2010, for a review). Given these findings, Trentacosta and Fine (2010) speculated that aggressive children and adolescents have a bias toward inferring anger when judging emotions and that this bias is more closely associated with aggressive behavior than is a general ERA deficit. This speculation also may apply to understanding the relationship between ERA and adult aggression. That is, although many studies have not found overall ERA differences between aggressive and nonaggressive adults (e.g., Best, Williams, & Coccaro, 2002; Larkin, Martin, & McClain, 2002), studies have found that aggressive, compared to nonaggressive, adults have a negativity bias and, in some cases an anger bias, when they make emotion recognition errors. For example, when emotion recognition errors are made, hostile and

This article was published Online First April 7, 2014.

Michael F. Wagner, Joel S. Milner, Randy J. McCarthy, Julie L. Crouch, Thomas R. McCanne, and John J. Skowronski, Department of Psychology, Center for the Study of Family Violence and Sexual Assault, Northern Illinois University.

Joel S. Milner is the author/copyright holder/royalty recipient for the Child Abuse Potential Inventory.

Correspondence concerning this article should be addressed to Joel S. Milner, Center for the Study of Family Violence and Sexual Assault, Northern Illinois University, 125 Presidents Blvd., DeKalb, IL 60115-2854. E-mail: jmilner@niu.edu

aggressive adults are more likely to label neutral faces as representing negative facial emotions (Best et al., 2002) and negative facial emotions as angry faces (Larkin et al., 2002), with some of these findings moderated by gender (Larkin et al., 2002).

With respect to parenting behaviors, it has been proposed that parental problems in ERA contribute to verbal and physical aggression toward children, including CPA. For example, the social information processing (SIP) model of CPA (Milner, 1993, 2000, 2003) suggests that not only are physically abusive parents and high-risk parents, compared to low-risk and nonabusive parents, less attentive to and less aware of their children's behavior, they also have difficulties accurately recognizing children's emotional states. Inaccurate recognition of a child's emotions coupled with the mislabeling of neutral emotions as negative emotions (as found in some studies investigating aggressive adults) might explain why physically abusive parents (Bauer & Twentyman, 1985) and highrisk for CPA parents (Farc, Crouch, Skowronski, & Milner, 2008) make more negative child-related attributions (e.g., hostile intent) when presented with neutral or ambiguous child cues, attributions that are associated with harsh parenting (MacKinnon-Lewis, Lamb, Arbuckle, Baradaran, & Volling, 1992; Slep & O'Leary, 1998). However, to date, only a few studies on the facial ERA of physically abusive parents and high-risk for CPA parents, relative to comparison parents, have been conducted, and the findings have been mixed.

In the first study of ERA in physically abusive parents, abusive mothers made more overall emotion recognition errors when asked to label seven infant facial emotions (Kropp & Haynes, 1987) than nonabusive mothers. When errors were made, abusive, compared to nonabusive, mothers tended to label positive faces as negative and significantly more often labeled negative faces as positive. In contrast, two subsequent studies reported that physically abusive, compared to nonabusive, mothers were not significantly different in their ability to identify children's facial emotions (Camras et al., 1988; During & McMahon, 1991) or adult facial emotions (During & McMahon, 1991); and, unlike the findings reported by Kropp and Haynes (1987), During and McMahon (1991) did not find an abusive parent negativity or positivity bias. However, in a study that evaluated perceptions of infant feelings (as opposed to ERA), Francis and Wolfe (2008) found that physically abusive, compared to nonabusive, fathers were more likely to perceive infants' emotional expressions as conveying a negative emotion, such as anger or disgust. No other published studies of ERA in physically abusive parents could be found. In summary, studies do not agree on the extent to which ERA differences (overall and with respect to a negativity bias) exist when physically abusive parents are compared to nonabusive parents.

In addition to studies on ERA in physically abusive and nonabusive parents, two studies have explored the extent to which highrisk for CPA, compared to low-risk for CPA, parents have emotion recognition difficulties. In the first study, high-risk mothers and low-risk mothers were asked to identify adult and child, facial and auditory emotions (i.e., angry, happy, fearful, sad) presented at low-intensity or high-intensity (Balge & Milner, 2000). Overall, high-risk, compared to low-risk, mothers showed a tendency to make more errors on the facial stimuli and on the low-intensity facial stimuli. ERA risk group differences only reached significance (p = .016) when a post hoc test of overall emotion recognition errors across all visual and auditory stimuli was conducted.

When emotion recognition errors were made, the types of emotion recognition errors were not reported. A second study on ERA in high-risk and low-risk mothers and fathers used two sets of facial emotions (set 1: angry, happy, fearful, sad, surprise, disgust, contempt; set 2: angry, happy, fearful, sad), each presented in a stress condition and a no stress condition (Asla, de Paúl, & Perez-Albéniz, 2011). Overall, on each set of facial stimuli, high-risk, compared to low-risk, parents made significantly (ps < .001) more emotion recognition errors. In addition, for both sets of facial emotions, there were gender effects that were qualified by a Risk Group X Gender interaction. High-risk fathers made more emotion recognition errors than high-risk mothers, whereas there were no differences in emotion recognition errors among low-risk fathers and low-risk mothers. Across the two sets of stimuli, the impact of stress was not consistent possibly due to the fact that parents did not perceive the stressor (i.e., crying child) as stressful.

Collectively, studies employing physically abusive or high-risk parents and comparison parents are variable with respect to whether or not they found ERA to be associated with CPA. It is unclear if the mixed findings are due to the relationship between ERA and CPA being weak (or nonexistent) or if a more substantial relationship could be found if the face recognition task was more challenging, such as when the "explicitness" of the facial expression is decreased. For example, Camras et al. (1988) suggested that more robust ERA differences between abusive and nonabusive parents may be present when the facial stimuli are viewed in less than optimal (real world) conditions, such as when facial emotions are viewed at an angle. This perspective is consistent with the aforementioned SIP model of CPA, which suggests that physically abusive parents and high-risk parents have difficulties accurately recognizing children's emotions (Milner, 2000); a deficit that could be due to emotion recognition skill deficits. We also speculated that limiting face presentation time might present a more challenging ERA task that might expose physically abusive parents' and high-risk parents' emotion recognition skill deficits. To date, however, no study has varied face angle and face presentation time when conducting ERA studies with high-risk or abusive parents.

Study 1: ERA in High-Risk Parents and Low-Risk Parents

To provide a facial emotion recognition task that varied in degree of difficulty, we presented high-risk parents and low-risk parents with child facial emotions (happy, fearful, angry, sad, neutral; which, except for neutral, were the emotions presented in each of the previous ERA studies using high-risk parents) that varied in presentation angle (full frontal, left 45 degrees, right 45 degrees, left 90 degrees, right 90 degrees) and presentation time (100 ms, 600 ms). A 100 ms presentation time was chosen as a viewing time close to recognition threshold; a 600 ms presentation time was chosen as a viewing time well above recognition threshold (assumptions supported by pilot data). Because gender (Mc-Clure, 2000), ethnic background (Elfenbein & Ambady, 2002), age (Ruffman, Henry, Livingstone, & Phillips, 2008), and education (Sasson et al., 2010) can impact facial ERA, across presentation conditions high-risk parents and low-risk parents were matched on gender, ethnic background, age, education, and to the extent possible on marital status and number of children. In addition, there is

some indication that a negativity bias exists when emotion recognition errors are made by physically abusive parents (Francis & Wolfe, 2008; Kropp & Haynes, 1987); therefore, we explored the extent to which a negativity bias characterized the emotion recognition errors made by high-risk parents.

We hypothesized that: (1) there would be ERA differences between presentation times (100 ms < 600 ms); emotions (with happy faces having the highest ERA; Elfenbein & Ambady, 2002; Langner et al., 2010), and presentation angles (90 degrees < 45 degrees < frontal views); (2) overall, high-risk parents would show a lower mean ERA than low-risk parents; (3) there will be interactions such that the largest ERA risk group differences would be found in the 90 degrees face presentation in the 100 ms presentation time condition; and (4) when emotion recognition errors were made, high-risk parents would more often label neutral faces as reflecting negative emotions (negativity bias) than low-risk parents.

Method

Participants and participant selection. Three hundred and 13 parents with normal or corrected-to-normal vision from communities surrounding a large university in the Midwestern United States participated in one of two facial emotion presentation time conditions: 100 ms (n=131) or 600 ms (n=182). The 313 parents served as a subject pool from which the final study participants were selected. For inclusion as high-risk parents, parents had to have valid Child Abuse Potential (CAP) Inventory abuse scores that were at or above the 215 abuse score cutoff (a cutoff score that determines scores in the outer 5% of the population; Milner, 1986). For inclusion as low-risk parents, parents had to have valid CAP abuse scores below 166 (an abuse cut-score based on signal detection theory, Milner, 1986). Among the 313 parents

in the general sample, 144 parents who had invalid CAP Inventories (i.e., elevated faking-good, faking-bad, or random-response indices) and 18 parents who had abuse scores between the 166–215 cut-scores on the CAP Inventory abuse scale were excluded. From the remaining pool of 151 high-risk parents and low-risk parents in the two presentation time conditions, parents were selected for inclusion in the study so that across risk groups and presentation times parents were matched on gender, ethnic background, age, and education.

The final study sample consisted of 30 low-risk parents (mean CAP abuse score = 83.0, SD=41.7) and 21 high-risk parents (mean CAP abuse score = 292.8, SD=38.6) in the 100 ms presentation time condition, and 37 low-risk parents (mean CAP abuse score = 100.2, SD=43.8) and 24 high-risk parents (mean CAP abuse score = 273.8, SD=34.0) in the 600 ms presentation time condition. As expected, a one-way analysis of variance (ANOVA) indicated a significant difference between the risk groups' abuse scores, F(3, 108) = 200.84, p < .001, $\eta_g^2 = .85$. A follow-up Student Newman Kuels (SNK) test indicated that the mean abuse scores for the two high-risk groups were not different (p = .09) and the mean abuse scores for the two low-risk groups were not different (p = .12), whereas each of the high-risk groups had higher (p < .001) abuse scores than each of the low-risk groups.

The demographic characteristics of the low-risk parents and high-risk parents across presentation time conditions are shown in Table 1. As expected, following matching, the four between groups conditions were not significantly different with respect to gender, $\chi^2(3) = 1.35$, p = .71; ethnic background, $\chi^2(3) = .53$, p = .91; age, F(3, 108) = 1.63, p = .19; education, F(3, 108) = 1.36, p = .26; and, number of children, F(3, 108) = 0.64, p = .59. However, groups were different with respect to marital status,

Table 1
Demographic Characteristics of Low-Risk Parents and High-Risk Parents Across Presentation
Time Conditions

	Presentation time condition $(N = 112)$					
	100 ms	(n = 51)	600 ms (n = 61)			
Characteristic	Low-risk $(n = 30)$	High-risk $(n = 21)$	Low-risk $(n = 37)$	High-risk $(n = 24)$		
Gender (%)						
Mothers	53.3	57.1	56.8	50.0		
Fathers	46.7	42.9	43.2	50.0		
Ethnic background (%)						
Caucasian	30.0	33.3	32.4	29.2		
African-American	70.0	66.7	67.6	70.8		
Age (years)						
M	38.0	31.6	33.8	34.5		
SD	12.7	9.6	8.7	11.3		
Education (years)						
M	13.4	12.6	12.9	12.8		
SD	1.8	1.5	1.6	1.3		
Marital status (%)						
Single/divorced	46.7	66.7	70.3	70.8		
Married/cohabitating	53.3	33.3	29.7	29.2		
Number of children						
M	2.2	2.1	2.6	2.5		
SD	1.9	1.3	1.3	1.6		

 $\chi^2(3) = 16.74$, p < .001. We computed the correlation between marital status (coded 0 = single/divorced, 1 = married/cohabiting) and ERA and found that the association was not significant, r(110) = -.13, p = .18. Therefore, we did not control for marital status in any of the ERA analyses.

Measures.

Personal data form. The personal data form requested general background information (e.g., gender, ethnicity, age, years of education, marital status, and number of children) from parents.

Child Abuse Potential (CAP) Inventory. The CAP Inventory is a 160-item, forced choice (agree/disagree) questionnaire designed to screen for child physical abuse risk (Milner, 1986, 1994, 2004). The questionnaire contains a physical abuse scale, six factor scales, and three validity scales (lie scale, random response scale, inconsistency scale), which form three response distortion indices (faking good, faking bad, random response indices). Internal consistency estimates for the abuse scale range from .92 to .96 for general population parents and from .95 to .98 for child physical abusers (Milner, 1986). In the present study, abuse scale internal consistency estimates were .91 and .94 for parents in the initial participant pool and in the final study sample.

Extensive construct validity data, indicating the expected relationships between risk factors and CAP abuse scores, have been reported (Milner, 1986, 1994, 2004). Concurrent predictive validity studies report that the CAP abuse scale correctly classifies 80% to 90% of child physical abusers and nonabusers (e.g., Milner, 1986, 1994). Future predictive validity data indicate that elevated abuse scores are significantly related to later incidents of CPA (Chaffin & Valle, 2003; Milner, Gold, Ayoub, & Jacewitz, 1984). In addition, prospective studies have reported that elevated CAP abuse scores are predictive of other negative child outcomes (e.g., neonatal morbidity [Zelenko et al., 2001], and children's future intelligence and adaptive behaviors [Dukewich, Borkowski, & Whitman, 1999]).

Experimental stimuli. The pictures of children's facial emotions used in the present study were taken from a set of standardized photographs—the Radboud Faces Database (RaFD, Langner et al., 2010)—which provides facial emotions presented in frontal and angle views. Specifically, the RaFD contains color portrait images of 49 models (39 Caucasian adults, 20 males and 19 females; 10 Caucasian children, 4 males and 6 females) expressing eight emotions (angry, fearful, happy, sad, neutral, surprise, disgust, contempt). Each emotion is displayed with the eyes directed straight ahead, averted to the left, or averted to the right. Photographs of each model displaying each of the eight emotions under each eye gaze condition were taken simultaneously from five viewpoints (full frontal, 45 degrees [left and right], and 90 degrees [left and right profile]). Facial emotions were based on prototypes described in the Facial Action Coding System (Ekman, Friesen, & Hager, 2002). Specific facial action units, adopted from the Directed Facial Action Task (Ekman, 2007), were used to develop each expression (Langner et al., 2010). Validation data indicate that the frontal images show a "very high recognition" of the intended emotion expression, with the highest accuracy rate (98%) for happiness and the lowest accuracy rate (50%) for contempt (Langner et al., 2010). The accuracy rates for the other six emotions were between 80% and 90%.

Four male child photographs and four female child photographs (the four female child photographs were chosen randomly from the

six available female photographs so that the numbers of male and female photographs were equal) were selected for use in the present study. The photographs depicted five facial emotions (angry, fearful, happy, sad, neutral) expressed with eye gaze directed straightforward in five views: frontal; left 45 degrees, right 45 degrees, left 90 degrees, right 90 degrees. Frontal views were presented twice so that two frontal views, two 45 degree views (left and right), and two 90 degree views (left and right) were shown. Thus, a total of 240 child faces (8 children \times 5 emotions \times 3 angles \times 2 presentations) were presented.

In addition, for use in practice trials, 10 photographs (the two female child models portraying each of the five emotions at the frontal view that were not used in the main study) constituted a pool from which five practice photographs were randomly selected with the restriction that each emotion was selected only once. The five practice photographs were presented in a different random order to each parent before the main part of the study began.

To determine if the computer software displayed facial expressions for the appropriate presentation times in the 100 ms and 600 ms conditions, a Thorlabs DET10A high-speed photo detector (placed in front of the computer screen) was used to record the presentation times of the frontal facial photographs in the 100 ms and 600 ms conditions. Across all emotions and all child models, presentation times for the 100 ms and 600 ms conditions were determined to be a mean of 118.7 ms (SD = 0.95) and a mean of 611.0 ms (SD = 1.01), respectively.

Procedures. All procedures were approved by the human subjects review board at the authors' university. After each participant read and signed an Informed Consent form, they were seated in front of a computer. Before beginning the emotion recognition task, participants were provided with instructions on the computer screen. They were told that a plus (+) sign would be presented at the center of the computer screen and that when they were ready to view a facial expression, they should focus on the plus sign and press the keyboard spacebar. Participants also were told that following each face presentation, they would be asked to indicate whether the expression was angry, happy, sad, fearful or neutral by pressing one of five computer keys that were labeled with one emotion word (angry, happy, sad, fearful or neutral).

After reading the instructions, participants completed five practice trials on which each of the five emotions was presented once in a frontal view for a presentation time that matched the presentation time condition to which the parent had been assigned. Next, participants viewed the 240 previously described photographs in a random order that was unique to each participant. Faces were presented in blocks of 20 trials with a short break between each block. Following completion of the emotion recognition task, parents completed the personal data form and the CAP Inventory. After completing the questionnaires, participants were again given the opportunity to ask questions, were thanked, and were given \$25.

Results and Discussion

Mean ERA proportions (correct hits out of total possible correct hits) for the different study conditions were analyzed using a four-factor ANOVA with repeated measures on the last two factors. The first two factors were risk group (low-risk, high-risk) and presentation time (100 ms, 600 ms); the last two factors were

emotion (angry, fearful, happy, sad, neutral) and presentation angle (frontal, 45 degrees, 90 degrees). Because the ERA proportions were not normally distributed, arcsine transformation of the proportion data was conducted before calculating the ANOVA. When the data violated the sphericity assumption (as indicated by Mauchly's test), the *ps* associated with effects were altered using the Greenhouse-Geisser adjustment; it is these corrected *p* values that are reported. Because all hypotheses were directional, one-tailed tests of significance were conducted. For ease of interpretation, the untransformed ERA means (*SDs*) for risk groups and presentation times are presented in Table 2.

As expected, the ANOVA yielded significant effects for face presentation time, $F(1, 108) = 42.18, p < .001, \eta_g^2 = .12$ (600) ms > 100 ms), face emotion $F(4, 432) = 183.93, p < .001, \eta_g^2 =$.39, and face presentation angle, F(2, 216) = 363.88, p < .001, $\eta_g^2 = .13$. SNK follow-up tests for the main effects for emotion and angle revealed the expected patterns of ERA means for angle (frontal > 45 degrees > 90 degrees) and for happy facial emotions (happy > fearful = neutral > angry = sad). However, the hypothesized risk group effect only approached significance, F(1, $108) = 2.26, p = .068, \eta_g^2 = .01$. Nevertheless, the means were as predicted with high-risk parents (M = .787, SD = .12) exhibiting a lower mean ERA than low-risk parents (M = .817, SD = .11). Contrary to expectations, risk group did not interact with any of the other study variables (ps > .05). Thus, although the variables that manipulated difficulty of the emotion recognition task were successful in impacting overall ERA, they did not differentially impact high-risk parents.

In a final analysis we tested the hypothesis that, when errors were made, high-risk, relative to low-risk, parents would more often label neutral faces as negative faces. However, there was no risk group difference in this type of emotion recognition confusion in either of the presentation time conditions (ps = .28 and .75 for the 100 ms and 600 ms presentation times conditions). We also conducted exploratory post hoc analyses to determine if, when errors were made, high-risk, relative to low-risk, parents would more often label neutral faces as angry faces and sad faces as angry faces in either of the presentation time conditions. These post hoc speculations were supported only in one case. When errors were made, there was a trend for high-risk, relative to low risk, parents to more often label the sad faces as angry faces in the 600 ms condition (p = .06, one-tailed test).

In summary, we investigated the extent to which a study that was designed to limit the facial information available for processing might enhance expected ERA differences between high-risk parents and low-risk parents. Although the manipulations used to limit facial information were successful, as indicated by reductions in ERA as face angle increased and face presentation time decreased, we only observed an overall tendency for high-risk, relative to low-risk, parents to display lower ERA; and risk group did not interact with any of the other study variables. These results (a nonsignificant trend) are similar to findings reported by Balge and Milner (2000). However, they contrast with a study by Asla et al. (2011) that reported high-risk, compared to low-risk, parents displayed significantly lower ERA. Although the present study and the studies by Balge and Milner and Asla et al. used the same risk measure (i.e., the CAP Inventory), one difference was that Asla et al. used a Spanish-speaking sample obtained in Spain, whereas the present study and the Balge and Milner study used Englishspeaking samples obtained in the United States. This suggests that future studies might examine the possibility of moderation of the CPA-risk/ERA relationship by culture.

Table 2
Mean (SD) ERA Proportions for Emotions Across Angles, Presentation Times, and Risk Groups

Emotion/Angle	Presentation time condition $(N = 112)$					
	100 ms	(n = 51)	600 ms (n = 61)			
	Low-risk $(n = 30)$	High-risk $(n = 21)$	Low-risk $(n = 37)$	High-risk $(n = 24)$		
Angry						
Frontal	.802 (.21)	.771 (.20)	.890 (.13)	.844 (.18)		
45 degrees	.621 (.23)	.604 (.20)	.807 (.15)	.771 (.18)		
90 degrees	.308 (.20)	.268 (.19)	.747 (.18)	.682 (.21)		
Fearful	` /	` /	` /	` '		
Frontal	.913 (.11)	.869 (.16)	.914 (.12)	.823 (.21)		
45 degrees	.892 (.11)	.851 (.14)	.929 (.11)	.878 (.19)		
90 degrees	.712 (.22)	.640 (.24)	.926 (.13)	.867 (.20)		
Нарру	` /	` /	` /	` '		
Frontal	.994 (.02)	.994 (.02)	.995 (.02)	.987 (.04)		
45 degrees	.992 (.02)	.994 (.02)	.998 (.01)	.984 (.04)		
90 degrees	.927 (.08)	.899 (.15)	.995 (.02)	.974 (.07)		
Sad	` /	` /	` /	` '		
Frontal	.750 (.17)	.821 (.18)	.856 (.15)	.818 (.17)		
45 degrees	.708 (.19)	.717 (.17)	.828 (.15)	.763 (.13)		
90 degrees	.281 (.13)	.336 (.19)	.623 (.16)	.594 (.16)		
Neutral	` /	` /	` /	` '		
Frontal	.775 (.24)	.774 (.24)	.892 (.13)	.836 (.23)		
45 degrees	.675 (.26)	.729 (.21)	.882 (.17)	.872 (.22)		
90 degrees	.748 (.24)	.702 (.22)	.924 (.11)	.846 (.24)		

Note. SD = standard deviation; ERA = emotion recognition accuracy.

With respect to the types of emotion recognition errors made, the prediction that high-CPA risk parents would more often label neutral faces as expressing negative emotions (negativity bias) than low CPA risk parents was not supported. Although a negativity bias has not been previously studied in parents who vary in CPA risk, additional data are needed to confirm this finding because it has occasionally been found in studies of parents who are physically abusive (e.g., Kropp & Haynes, 1987).

Finally, in an exploratory post hoc analysis, we found that when errors were made in the 600 ms presentation time condition, high-CPA-risk parents more often tended to label sad faces as angry faces than low-CPA-risk parents. If replicated, future research could explore the possibility that this confusion may contribute to parental aggression, perhaps by increasing the parent's attributions of a child's hostile intent. Further, because the mislabeling of sadness as anger might limit parents' perspective-taking ability (a component of empathy), the possibility that this confusion may contribute to the observation that high-risk, unlike low-risk, parents do not show increased empathy toward a crying infant (Milner, Halsey, & Fultz, 1995) could be explored.

Study 2: Meta-Analysis of ERA in High-Risk, Physically Abusive, and Comparison Parents

The relationship between ERA and CPA is not as definitive as one would like. Certainly the weakness of some of the results for studies that have investigated ERA differences between high-risk for CPA or abusive parents and comparison parents leaves some doubt as to whether such a relationship exists. In response to such doubt, we conducted a meta-analysis. This meta-analysis addressed two research questions: (1) collectively, do published studies show that high-risk and physically abusive parents, relative to comparison parents, exhibit lower ERA for child facial emotions?; and, (2) are any observed ERA effects for child facial emotions moderated by the population studied (high-risk parents vs. physically abusive parents)?

Method

Literature search. To locate studies that explored the extent to which ERA varied across parents who differed in CPA risk or differed across physically abusive parents and nonabusers, we used various relevant terms (e.g., child abuse, child physical abuse, child maltreatment, high-risk parents, face emotions, emotion recognition, emotion recognition accuracy) to search a variety of databases (e.g., ArticleFirst, ERIC, MEDLINE, PsycArticles, PsycInfo, PubMed, Social Services Abstracts, Sociological abstracts, Web of Science, and WorldCat) and Internet search engines (i.e., Excite, Yahoo, Google, and Bing).

Inclusion criteria. Studies were included in the meta-analysis if they: (a) studied ERA in CPA groups (i.e., high-risk or physically abusive parents) and comparison groups; (b) measured facial ERA using procedures that asked parents to label children's facial emotions depicted in photographs or electronic images; and, (c) were published in a peer-reviewed journal. Only five studies met the criteria for inclusion in the meta-analysis. Two studies compared high-risk parents with low-risk parents (Asla et al., 2011; Balge & Milner, 2000), and three studies compared physically abusive parents with nonabusive parents (Camras et al., 1988;

During & McMahon, 1991; Kropp & Haynes, 1987). In addition, ERA data from the high-risk parents and low-risk parents in the present study (Study 1) were included in the meta-analysis (see Table 3 for a description of each study).

Data abstraction. A data abstraction form was used to summarize the distinguishing features of each study (e.g., number and types of groups contrasted, study design, description of the emotions evaluated, etc.) and the statistical data available for use in the meta-analysis. The first and second authors independently abstracted information from studies. After all article reviews were completed, data abstractions results were compared and discrepancies were resolved.

Effect size calculation. Cohen's d, the difference between groups divided by the pooled standard deviation, was used as the measure of ES (Lipsey & Wilson, 2001). One study (Camras et al., 1988) did not report sufficient information required for computing d. In an attempt to obtain the needed information, we contacted the senior author of the study who informed us that the information needed to compute d was no longer available (L. A. Camras, personal communication, October 31, 2012). Because Camras et al. reported that abusive, relative to the nonabusive, parents did not differ in ERA, for the purpose of including their data in the meta-analysis we set the d to zero to reflect their null results.

Results and Discussion

A random-effects model was used for the meta-analysis (Hedges & Vevea, 1998). The overall effect size was a weighted mean effect size. It would have been desirable also to adjust effect sizes for the reliability of the facial expressions. Unfortunately, the needed reliability data are not available for each of the facial expressions. However, it should be pointed out that making this adjustment in effect sizes improves the effect size estimate by "accounting for the measurement error attenuating each estimate" (Ellis, 2010, p. 103). Thus, the consequences of our not making this adjustment means that our estimate of the overall effect size is a conservative or underestimate of the effect size.

To test for heterogeneity in ESs among studies, we computed *Q*. In addition, we explored whether the ES varied by the population studied by comparing high-risk parents' and abusive parents' ESs. To assess for publication bias, we examined the correlation between study sample size and study ES (Lipsey & Wilson, 2001). Finally, we calculated a fail-safe *N*, an index of the number of studies required to reduce the weighted mean ES to a nonsignificant level (Orwin, 1983).

Results of the meta-analysis (see Table 3) revealed that overall the combined high-risk and abusive parent groups, relative to their comparison groups, exhibited significantly lower ERA, weighted mean d=0.467, 95% CI [0.197, 0.738], k=7, N=414. However, significant heterogeneity in the study ds was found, Q=70.5, p<.001. To determine if the heterogeneity was due to the population studied, we compared the weighted mean d for studies that used high-risk parents and studies that used physically abusive parents. Analysis revealed that the difference in ESs for these two groups of studies was not significant, as indicated by overlapping CIs for the weighted mean ds (studies using high-risk parents: d=0.486, 95% CI [0.105, 0.868], k=4, N=288; studies using abusive parents: d=0.447, 95% CI [-0.043, 0.936], k=3, N=126). To assess for publication bias, we determined the extent that

Table 3
Studies Included in the Meta-Analyses of Child Facial ERA in High-Risk, Physically Abusive, and Comparison Parent Groups

	$\frac{\text{High-risk}}{n}$	$\frac{\text{Low-risk}}{n}$	$\frac{\text{Total}}{N}$	Child facial emotions	d^{b}	Q
				evaluated		
Study (High-risk/low-risk parents) ^a						
Asla et al. (2011) ^c fathers	24	40	64	Angry, happy, fearful, sad	1.280	
Asla et al. (2011) ^c mothers	40	40	80	Angry, happy, fearful, sad	0.310	
Balge & Milner (2000) ^d	16	16	32	Angry, happy, fearful, sad	0.148	
Study 1 ^e	45	67	112	Angry, happy, fearful, sad,	0.260	
				neutral		
Total N/Effect size	125	163	288		0.486 ^g	46.7***
	Abusers	Nonabusers	Total			
Study (Abusive/nonabusive parents)	n	n	N			
Camras et al. (1988)	20	20	40	Angry, happy, fearful, sad, disgust, surprise	$0.000^{\rm f}$	
During & McMahon (1991)	23	23	46	Angry, happy, fearful, sad, disgust, surprise	0.457	
Kropp & Haynes (1987)	20	20	40	Angry, joy, fearful, sad, surprise, interest, distress/pain	0.907	
Total N/Effect size	63	63	126	r	0.447^{g}	24.1***
Grand total N/Effect size	188	226	414		0.467^{g^*}	70.7***

Note. ERA = emotion recognition accuracy.

^a All risk groups were determined using the Child Abuse Potential Inventory. ^b Positive ds reflect a facial ERA advantage for the low-risk/nonabusive parents. ^c Two sets of facial stimuli were presented; however, only data from the Diagnostic Analysis of Nonverbal Behavior 2, (the measure used in the other two high- versus low-risk parent studies) were used in the meta-analysis; and, there was a risk by gender interaction, so effect sizes are presented by gender. ^d Visual and auditory child and adult stimuli were studied; however, only child facial data were used in this meta-analysis. ^e Study 1 refers to Study 1 in the present report. Further, although fathers and mothers were included in this study, results could not be broken out by gender because mothers and fathers could not be matched on demographics and the resulting sample sizes would be small. ^f Camras et al. (1988) did not report sufficient information to compute d; to reflect their null results, we set the d to zero. ^g Weighted mean d. ^{*} p < .05. *** p < .001.

ESs varied as a function of sample size; the correlation was not significant, r(5) = -.01, p = .98. In addition, the fail-safe N, the number of unpublished (file drawer) studies needed to reduce the current ES (d = 0.467) to a negligible level (d = 0.100), was estimated to be 25.7 studies.

Thus, despite the appearance of mixed findings that might be provided by a narrative overview of the literature, the metaanalysis revealed that across high-risk and physically abusive parent groups, relative to comparison groups, not only was there an ERA difference (i.e., d of 0.467) but it was a medium ES (using Lipsey's [1990], description of ds less than 0.320 as "small," ds between 0.330 and 0.550 as "medium," and ds greater than 0.560 as "large"). Moreover, the mean ESs exhibited by these two different populations were similar, so the significant heterogeneity across study ds cannot be attributed to whether a study examined high-risk for CPA parents or physical child abusers. This latter finding is consistent with the view that ERA deficits precede the CPA event and are not the result of parent-child abusive interactions. However, to document fully this possibility, a study of high-risk and low-risk individuals who have not yet had children needs to be undertaken.

Because the ES difference between the target populations (abusers vs. high-risk parents) did not contribute to the large heterogeneity in *ds* found across studies, other moderators need to be considered (e.g., criteria used to define risk, CPA, and comparison

groups; rigor of experimental design, the possible effect of gender). Unfortunately, the small number of studies in the meta-analysis precluded the possibility of a search for additional moderators. Nevertheless, to test the possible importance of group definitions, we conducted a post hoc analysis of Study 1 data where we redefined low-risk parents as those parents having a CAP abuse score of 91 (the CAP abuse score mean [Milner, 1986]) or less, instead of using abuse scores below 166 score (the CAP abuse cut-score based on signal detection theory, Milner, 1986).

A more conservative abuse score was selected to reduce the likelihood of false negative classifications in the low-risk group. In this analysis, where high-risk and low-risk parents were again demographically matched, the risk Group ERA difference was significant, F(1, 72) = 5.16, p = .013, $\eta_g^2 = .03$. This significant finding contrasts with the finding of only a trend in the direction of a risk Group ERA difference in the full Study 1 sample. Notably, this effect was found despite the substantial reduction (removal of 36 parents) in the low-risk group. In addition, since the metaanalysis included an independent sample of fathers and this sample had the largest effect size (see Asla et al., 2011 study in Table 3), for informational purposes we conducted a second post hoc analysis where this father group was removed from the overall metaanalysis. The reanalysis revealed an overall weighted mean of d =.321, 95% CI [0.111, 0.532], k = 6, N = 350; where p = .003. This analysis indicated that, although the effect size decreased, the overall effect remained significant after removing the results for the group that contained only fathers.

With respect to the presence of publication bias, the nonsignificant correlation between study sample sizes and ESs suggest an absence of publication bias (which is often presented visually in a funnel plot) based on the assumption that if no publication bias exists the study ESs should be distributed equally across sample sizes (Lipsey & Wilson, 2001). When we assessed for the threat of publication bias by computing a fail-safe N, we estimated that 25.7 studies with an effect size of zero (d=0.00) and average sample size would be needed to reduce the present ES to a minimal (i.e., d=0.100) level. Although not conclusive, these findings lend support to the view that the outcome of the present meta-analysis, which indicated a medium (ES) difference in ERA between CPA groups, is reliable.

General Discussion

The data described in this article support the contention that parents at high risk for CPA and child physical abusers, relative to comparison parents, make more child facial emotion recognition errors.

Additional Limitations and Future Research

The present research study (Study 1) and the meta-analysis (Study 2) have a number of additional limitations. For example, it cannot be assumed that the ERA in Study 1 and most of the other studies reviewed in the meta-analysis reflect the accuracy with which parents can identify spontaneous child facial emotions. This is because, except for one study (Kropp & Haynes, 1987), the children in the photographs were modeling facial emotions as opposed to displaying spontaneous emotions. Another limitation is that all studies reviewed assessed ERA for static, not dynamic, facial expressions. Thus, research is needed to test whether the results reported here are generalizable to spontaneous and to dynamic child facial emotions. Further, the present findings are limited because in all of the studies reviewed, parental responses to an unknown child were assessed. Maternal responses to facial emotions of their own child versus the facial emotions of other child faces have been shown to differ, at least in nonabusive mothers (Strathearn, Li, Fonagy, & Montague, 2008).

Future studies also might explore ERA differences in high-risk/ abusive, relative to comparison, parents when contextual factors are present. For example, the SIP model of CPA suggests that when parents are assessing children's emotions in the presence of stressors, high-risk and abusive parents will exhibit lower ERA because they are more reactive to stress (Milner, 2000). Although two of the reviewed studies (Asla et al., 2011; Balge & Milner, 2000) tested this hypothesis and found inconsistent results with respect to the risk-related impact of a situational stressor (i.e., a crying infant) on ERA, it is noteworthy that parents in both studies did not perceive the stressor as stressful. Therefore, additional research is needed to determine if stressors perceived by parents as distressing have a larger impact on ERA in high-risk/abusive parents than in comparison parents. In addition, mechanisms that might mediate the expected effect of stress on ERA differences in high-risk/abusive parents might be investigated. Further, to the extent that ERA reflects the cognitive effort, investigators could

manipulate parents' cognitive loads (Baddeley, 1996) or deplete parents' self-regulatory resources (Baumeister, Muraven, & Tice, 2000); manipulations that might differentially impact high-risk and abusive, relative to comparison, parents because high-risk and abusive parents are believed to have limited cognitive and emotion regulation resources (Milner, 1994, 2000). Finally, the present research does not address the question of the extent to which parents' emotion recognition deficits must be combined with other risk factors (i.e., parenting skill deficits) for CPA to occur.

Clinical Implications

Although the mechanisms through which emotion recognition errors impact parental behavior as well as children's development remain to be explained, the results reported in the present article suggest that professionals should assess for ERA deficits when working with high-risk and abusive parents as well as potential parents. This recommendation is appropriate because, regardless of the exact nature and degree of the relationship between ERA and CPA, any degree of parental ERA deficit can potentially contribute to an array of problematic parent—child interactions due to the disruption in the parent's ability to understand the child's emotional state.

References

References marked with an asterisk indicate studies included in the meta-analysis.

*Asla, N., de Paúl, J., & Pérez-Albéniz, A. (2011). Emotion recognition in fathers and mothers at high-risk for child physical abuse. *Child Abuse & Neglect*, 35, 712–721. doi:10.1016/j.chiabu.2011.05.010

Baddeley, A. (1996). The fractionation of working memory. *Proceedings of the National Academy of Sciences of the United States of America*, 93, 13468–13472. doi:10.1073/pnas.93.24.13468

*Balge, K. A., & Milner, J. S. (2000). Emotion recognition ability in mothers at high and low risk for child physical abuse. *Child Abuse & Neglect*, 24, 1289–1298. doi:10.1016/S0145-2134(00)00188-5

Bauer, W. D., & Twentyman, C. T. (1985). Abusing, neglectful, and comparison mothers' responses to child-related and non-child-related stressors. *Journal of Consulting and Clinical Psychology*, 53, 335–343. doi:10.1037/0022-006X.53.3.335

Baumeister, R. F., Muraven, M., & Tice, D. M. (2000). Ego depletion: A resource model of volition, self-regulation, and controlled processing. *Social Cognition*, 18, 130–150. doi:10.1521/soco.2000.18.2.130

Best, M., Williams, J. M., & Coccaro, E. F. (2002). Evidence for a dysfunctional prefrontal circuit in patients with an impulsive aggressive disorder. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 8448–8453. doi:10.1073/pnas.112604099

*Camras, L. A., Ribordy, S., Hill, J., Martino, S., Spaccarelli, S., & Stefani, R. (1988). Recognition and posing of emotional expressions by abused children and their mothers. *Developmental Psychology*, 24, 776–781. doi:10.1037/0012-1649.24.6.776

Chaffin, M., & Valle, L. A. (2003). Dynamic prediction characteristics of the Child Abuse Potential Inventory. *Child Abuse & Neglect*, 27, 463– 481. doi:10.1016/S0145-2134(03)00036-X

Dix, T. (1991). The affective organization of parenting: Adaptive and maladaptive processes. *Psychological Bulletin*, 110, 3–25. doi:10.1037/ 0033-2909.110.1.3

Dukewich, T. L., Borkowski, J. G., & Whitman, T. L. (1999). A longitudinal analysis of maternal abuse potential and development delays in children of adolescent mothers. *Child Abuse & Neglect*, 23, 405–420. doi:10.1016/S0145-2134(99)00019-8

*During, S. M., & McMahon, R. J. (1991). Recognition of emotional facial expressions by abusive mothers and their children. *Journal of Clinical Child Psychology*, 20, 132–139. doi:10.1207/s15374424jccp2002_4

- Ekman, P. (2007). The directed facial action task: Emotional responses without appraisal. In J. A. Coan & J. J. B. Allen (Eds.), *Handbook of emotion elicitation and assessment* (pp. 47–53). New York, NY: Oxford University Press
- Ekman, P., Friesen, W. V., & Hager, J. C. (Eds.). (2002). Facial action coding system (E-book). Salt Lake City, UT: Research Nexus.
- Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, 128, 203–235. doi:10.1037/0033-2909.128.2.203
- Ellis, P. D. (2010). The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results. New York, NY: Cambridge University Press.
- Farc, M. M., Crouch, J. L., Skowronski, J. J., & Milner, J. S. (2008). Hostility ratings by parents at risk for child abuse: Impact of chronic and temporary schema activation. *Child Abuse & Neglect*, 32, 177–193. doi:10.1016/j.chiabu.2007.06.001
- Francis, K. J., & Wolfe, D. A. (2008). Cognitive and emotional differences between abusive and non-abusive fathers. *Child Abuse & Neglect*, *32*, 1127–1137. doi:10.1016/j.chiabu.2008.05.007
- Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods*, 3, 486–504. doi:10.1037/1082-989X.3.4.486
- Izard, C. (1971). The face of emotion. East Norwalk, CT: Appleton-Century-Crofts.
- *Kropp, J. P., & Haynes, O. M. (1987). Abusive and nonabusive mothers' ability to identify general and specific emotion signals of infants. *Child Development*, 58, 187–190. doi:10.2307/1130300
- Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. F. M. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and Emotion*, 24, 1377–1388. doi: 10.1080/02699930903485076
- Larkin, K. T., Martin, R. R., & McClain, S. E. (2002). Cynical hostility and the accuracy of decoding facial expressions of emotions. *Journal of Behavioral Medicine*, 25, 285–292. doi:10.1023/A:1015384812283
- Lipsey, M. W. (1990). Design sensitivity: Statistical power for experimental research. Newbury Park, CA: Sage.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
- MacKinnon-Lewis, C., Lamb, M. E., Arbuckle, B., Baradaran, L. P., & Volling, B. L. (1992). The relationship between biased maternal and filial attributions and the aggressiveness of their interactions. *Development and Psychopathology*, 4, 403–415. doi:10.1017/S0954579400000869
- Mayer, J. D., Roberts, R. D., & Barsade, S. G. (2008). Human abilities: Emotional intelligence. *Annual Review of Psychology*, 59, 507–536. doi:10.1146/annurev.psych.59.103006.093646
- McClure, E. B. (2000). A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychological Bulletin*, 126, 424–453. doi:10.1037/0033-2909.126.3.424
- Milner, J. S. (1986). *The Child Abuse Potential Inventory: Manual* (2nd ed.). Webster, NC: Psytec.
- Milner, J. S. (1993). Social information processing and physical child abuse. *Clinical Psychology Review*, 13, 275–294. doi:10.1016/0272-7358(93)90024-G

- Milner, J. S. (1994). Assessing physical child abuse risk: The Child Abuse Potential Inventory. *Clinical Psychology Review*, 14, 547–583. doi: 10.1016/0272-7358(94)90017-5
- Milner, J. S. (2000). Social information processing and child physical abuse: Theory and research. In D. J. Hansen (Ed.), *Nebraska Symposium* on *Motivation*, *Vol. 45*, *Motivation and child maltreatment* (pp. 39–84). Lincoln, NE: University of Nebraska Press.
- Milner, J. S. (2003). Social information processing in high-risk and physically abusive parents. *Child Abuse & Neglect*, 27, 7–20. doi:10.1016/S0145-2134(02)00506-9
- Milner, J. S. (2004). The Child Abuse Potential (CAP) Inventory. In M. J. Hilsenroth & D. L. Segal (Eds.), *Comprehensive handbook of psychological assessment, Vol. 2: Personality assessment* (pp. 237–246). Hoboken, NJ: Wiley.
- Milner, J. S., Gold, R. G., Ayoub, C. A., & Jacewitz, M. M. (1984).
 Predictive validity of the Child Abuse Potential Inventory. *Journal of Consulting and Clinical Psychology*, 52, 879–884. doi:10.1037/0022-006X.52.5.879
- Milner, J. S., Halsey, L. S., & Fultz, J. (1995). Empathic responsiveness and affective reactivity in infant stimuli in high- and low-risk for physical child abuse mothers. *Child Abuse & Neglect*, 19, 767–780. doi:10.1016/0145-2134(95)00035-7
- Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. Journal of Educational Statistics, 8, 157–159. doi:10.2307/1164923
- Ruffman, T., Henry, J. D., Livingstone, V., & Phillips, L. H. (2008). A meta-analytic review of emotion recognition and aging: Implications for neuropsychological models of aging. *Neuroscience and Biobehavioral Reviews*, 32, 863–881. doi:10.1016/j.neubiorev.2008.01.001
- Saarni, C. (1999). The development of emotional competence. New York, NY: Guilford Press.
- Sasson, N. J., Pinkham, A. E., Richard, J., Hughett, P., Gur, R. E., & Gur, R. C. (2010). Controlling for response biases clarifies sex and age differences in facial affect recognition. *Journal of Nonverbal Behavior*, 34, 207–221. doi:10.1007/s10919-010-0092-z
- Shields, A., & Cicchetti, D. (1998). Reactive aggression among maltreated children: The contributions of attention and emotion dysregulation. *Journal of Clinical Child Psychology*, 27, 381–395. doi:10.1207/ s15374424jccp2704_2
- Slep, A. M. S., & O'Leary, S. (1998). The effects of maternal attributions on parenting: An experimental analysis. *Journal of Family Psychology*, 12, 234–243. doi:10.1037/0893-3200.12.2.234
- Strathearn, L., Li, J., Fonagy, P., & Montague, P. R. (2008). What's in a smile? Maternal brain responses to infant facial cues. *Pediatrics*, 122, 40–51. doi:10.1542/peds.2007-1566
- Trentacosta, C. J., & Fine, S. E. (2010). Emotional knowledge, social competence, and behavioral problems in childhood and adolescence: A meta-analytic review. *Social Development*, 19, 1–29. doi:10.1111/j.1467-9507.2009.00543.x
- Zelenko, M. A., Huffman, L. C., Brown, B. W., Jr., Daniels, K., Lock, J., Kennedy, Q., & Steiner, H. (2001). The Child Abuse Potential Inventory and pregnancy outcome in expectant adolescent mothers. *Child Abuse & Neglect*, 25, 1481–1495. doi:10.1016/S0145-2134(01)00284-8

Received June 16, 2013
Revision received January 6, 2014
Accepted January 17, 2014