Evolutionary Psychological Science
https://doi.org/10.1007/s40806-020-00254-w

RESEARCH ARTICLE

Sex Differences in the Recognition of Children's Emotional Expressions: A Test of the Fitness Threat Hypothesis

Elizabeth Hampson[1,2] · Paul Istasy[2] · Sawayra Owais[2] · Jessica A. Chow[1] · Belal Howidi[1] · Sarah J. Ouellette[1]

Received: 27 April 2020 / Revised: 30 June 2020 / Accepted: 1 July 2020

Abstract
Evolutionary theories have suggested that a female superiority in the recognition of facial emotion may be an adaptation that arises from women's greater responsibility and investment in child-rearing and infant care. In a previous study, we showed a female superiority on a set of computer-administered emotion recognition tasks that was most prominent for the discrimination of negatively as opposed to positively valenced facial expressions (e.g., fear), providing empirical support for the "fitness threat" hypothesis. In the present study, we further investigated sex differences in a new sample of 95 healthy men and women of reproductive age ($M_{age} = 22.09$ years), using images of both children's and adult's faces as stimuli to evaluate the speed and accuracy of emotion recognition. A female superiority in accuracy, which was more pronounced for negative than positive expressions, was found for adult face stimuli, replicating our previous findings. The sex difference was shown to extend robustly to infant and toddler faces, which represent a more ecologically valid test of the fitness threat hypothesis. Direct parenting experience, but not other forms of learned experience involving young children, was also found to be associated with the accuracy of emotion discrimination. Implications of this association are discussed.

Keywords Sex difference · Gender difference · Facial expression · Emotion · Fitness threat

Introduction

The capacity to decode facial expressions of emotion is important to everyday social interaction and is diminished in certain developmental disorders (e.g., Van Rheenen and Rossell 2014; Kuharic et al. 2019) or in individuals who experienced early life adversity (Russo et al. 2015). The latter observation suggests that accurate facial decoding is at least partly learned. However, evolved mechanisms are also believed to exist, which enable a receiver to judge another individual's emotional state on the basis of a characteristic facial signal and thereby anticipate future actions (Ekman 1997; c.f., Russell et al. 2003). Ekman and others (Ekman 1994; Izard 1994) proposed that a limited set of facial expressions is innate and carries universal signal value for the emotions happiness, sadness, anger, fear, disgust, and surprise. Parallel expressions are observed in other primate species (Van Hooff 1976). In humans, the production of this basic set of emotional expressions and their interpretation by a perceiver is found cross-culturally (Ekman et al. 1987; but see Russell et al. 2003).

Studies of the basic set of emotions have often, though not invariably, reported a sex difference, which can be detected as early as infancy (McClure 2000) but is largest in young adults (Thompson and Voyer 2014) and varies in magnitude depending upon the methodology used to evaluate it. Women and girls reportedly enjoy a modest advantage over men and boys in the rapid and/or accurate discrimination of facial expressions (e.g., Hampson et al. 2006; Anderson et al. 2011; for a review, see Thompson and Voyer 2014). However, the sex difference ranges from small to moderate and is not always found (e.g., Calder et al. 2003; Grimshaw et al. 2004). Some researchers have argued that any genuine sex difference might be limited to a single emotion (e.g., disgust, Connolly et al. 2019; or sadness, Mandal and Palchoudhury 1985), or may be dependent on subject variables such as age of the perceiver (Williams et al. 2009). A recent meta-analysis of 551 effect sizes from 215 samples concluded that a sex difference in favor of females does exist, but the magnitude of the female

✉ Elizabeth Hampson
ehampson@uwo.ca

1 Department of Psychology, University of Western Ontario, London, ON N6A 5C2, Canada

2 Program in Neuroscience, University of Western Ontario, London, Ontario N6A 5C2, Canada

Published online: 09 July 2020

superiority appears to be moderated by a number of factors, many of them not completely understood (Thompson and Voyer 2014).

In a previous study (Hampson et al. 2006), we found that the emotional valence of a facial display may be one important moderator of whether or not a sex difference is seen, and its magnitude. Using a set of computerized forced-choice tasks, we found that women were faster than men at recognizing both positive and negative emotions from facial cues, but the sex difference was accentuated for negatively valenced emotions. This interaction was interpreted as preliminary support for the "fitness threat hypothesis" (described in Hampson et al. 2006). The fitness threat hypothesis derives from a broader theory by Babchuk and colleagues ("the primary caretaker hypothesis"; Babchuk et al. 1985), which argues that females, as the sex that dominated childcare throughout hominid evolution, have evolved specific adaptations that increase the probability of offspring survival (Babchuk et al. 1985). The fitness threat hypothesis elaborates on this idea by emphasizing the special significance of negative emotions. It predicts that a female superiority in emotional decoding may be limited to negative expressions because it is specifically negative expressions, when displayed by infants or toddlers, that signal a potential threat to offspring survival (e.g., fear may denote the presence of danger, disgust the ingestion of a noxious or harmful substance, or sadness either physical pain or a failure to thrive), calling for action on the caretaker's part. The fitness threat hypothesis does not deny that these emotions are separate entities, both neurologically and functionally, yet they are united by the fact that they call for intervention by a caretaker (Hampson et al. 2006). On the other hand, positive emotional displays do not carry any immediate survival imperative. Although both sexes do have a stake in infant survival, the prompt and accurate identification of a potential threat to well-being is a fundamental adaptation to the role of primary caretaker and would be maximized in the sex making the largest investment in offspring.

The fitness threat hypothesis has received preliminary support in our own and subsequent studies (e.g., Anderson et al. 2011; Hampson et al. 2006; Parsons et al. 2019; Williams et al. 2009; see Thompson and Voyer 2014, for a review), but explicit and deliberate tests of the hypothesis are still lacking. Importantly, it has yet to be tested by using children's faces as visual stimuli. Progress has been hindered by the absence of well-validated, standardized image databases consisting of infants' and toddlers' faces that depict the full spectrum of basic emotions (but see Maack et al. 2017). For this reason, nearly all the existing research has focused on sex differences in the recognition of emotion in *adult* faces.

One exception is Babchuk et al. (1985) who reported an overall female advantage in the verbal labelling of children's emotions, using as stimuli 20 slides of children's faces depicting 8 identifiable emotions (e.g., fear, physical distress). A main effect of sex collapsed across the 8 emotions was reported. However, the magnitude of the sex difference was not compared for positive versus negative expressions. More recently, Proverbio and colleagues presented infant facial expressions chosen to be either weakly or strongly positive or negative and found that women were more accurate than men at interpreting the directional valence of the expressions, "especially weakly negative expressions" (Proverbio et al. 2007, p. 480). Though not expressly designed to test the fitness threat hypothesis, Proverbio et al.'s study supports the possibility that negative faces elicit a larger female advantage.

Relatively little work has explored the possible role of expertise or experience in modifying the ability to discriminate facial emotion. The decoding of children's faces brings up the issue of expertise effects directly because, unlike judgments based on adult faces, men and women may differ substantially in their cumulative past experience with infants and toddlers. A role for learning in facial affect decoding has been suggested (Tottenham 2013), acting in conjunction with neurobiological mechanisms thought to be innate. Learning processes may contribute to an individual's ability to assign emotional meaning to facial expressions. Such learning may be acquired beginning in childhood via classical conditioning and exposure to specific facial configurations that typify the human species in their natural emotional context. Thus, there may be plasticity in the ability to decode faces, which is contingent upon experience. A role for experience is supported by alterations in facial processing reported to result from social deprivation in childhood or by improvements in facial decoding in special populations after deliberate instruction and practice (Javanbakht et al. 2015; Lopata et al. 2012; Russo et al. 2015; Székely et al. 2014). However, whether expertise acquired at older ages (e.g., beyond childhood) can refine the speed or accuracy of facial decoding is unclear.

Neither Babchuk et al. (1985) nor Hampson et al. (2006) found that direct experience with young children played a significant role in the sex differences they observed. However, parenting one's *own* children confers intensive experience and might occupy a special status. In favor of parenting having a potential to modify the decoding of children's faces, infants' and children's faces elicit greater attentional engagement in adult observers than adult faces do, particularly among those with parenting experience (Thompson-Booth et al. 2014). Furthermore, brain regions influenced by emotion such as the prefrontal cortex may be differentially activated in parents versus non-parents by viewing emotional infant faces (Nishitani et al. 2011; Proverbio et al. 2006). In addition, Proverbio et al. (2007) found that "expert" viewers (9 women, 8 men), some of whom were parents, were more accurate than non-experts in identifying the emotional valence (positive or negative) and emotional intensity (weak or strong) of infant faces. Whether the *sex difference* in recognizing facial emotion is influenced by parental experience has not been established. In fact, this important demographic characteristic has been largely overlooked by most researchers. To the extent that accuracy of emotion recognition is influenced by experiential and not just innate variables, it is imperative that researchers begin to inquire about parental experience, in order to determine if any sex difference in the recognition of children's emotional expressions is moderated, or even explained, by differences in accumulated past exposure to young children.

The present study was designed to further test the fitness threat theory proposed by Hampson et al. (2006). Both adult faces and, importantly, children's faces were employed as stimuli. Three separate statistical hypotheses derived from the fitness threat theory were identified a priori and tested using a set of planned *F* tests. Each hypothesis predicted the presence of a specific valence-related interaction effect which, if supported, would constitute further evidence for the fitness threat proposal:

1. We hypothesized that a sex difference would be found in the recognition of emotion in adult faces and would be most prominent for negative emotional expressions. Specifically, a female advantage that is selective to negative emotions would support the fitness threat hypothesis.
2. The fitness threat hypothesis proposes that the female advantage derives from the necessity to properly decode emotion in the facial expressions of young dependent children who have limited verbal skills. Accordingly, we predicted that a female advantage, larger for negative than positive expressions, would be found for infants' and toddler's faces. This interaction effect is at the heart of the fitness threat hypothesis.
3. Because of their greater ecological validity, we further predicted that the magnitude of the sex difference might be enhanced for children's faces relative to adult faces. To test for a difference in effect size, responses to the child and adult faces were contrasted directly in a third analysis. A three-way interaction between sex of observer, emotional valence, and age of the stimulus face was expected.

To enable the clearest test of the fitness threat hypothesis, we kept track of parental status among our study volunteers. The three hypotheses were evaluated conservatively, with only *non*-parents included in all statistical analyses. In keeping with the concept of an evolved adaptation, we predicted that parental experience would *not* be a necessary prerequisite for a sex difference to be seen. In addition, because our study recruitment did capture a small number of individuals with direct parenting experience ($n = 23$), we performed an exploratory data check to evaluate whether greater experience with young children acquired through parenting was associated with increased accuracy in the decoding of children's facial expressions.

Method

Participants

Ninety-five volunteers (50 women, 45 men) participated. Target sample size was established based on our previous work that used the same emotional recognition task as the present study and its estimated effect size of $f = .25$ (Hampson et al. 2006). Seventy-two were non-parents and 23 were parents of a child aged 5 or younger. The mean age was 22.09 ± 6.78 years. Parents reported having 1–2 children on average ($M = 1.57$). Participants were recruited through flyers or an on-line portal at Western University and participated for a research course credit or were reimbursed for their participation. All participants had normal or corrected vision.

Procedure

Testing was conducted individually. The session began with a Snellen visual acuity test followed by a standardized mood questionnaire (Profile of Mood States; McNair et al. 1971). Participants then completed a set of face discrimination tasks (described below) that began with a practice condition (Face Matching) and ended with a non-face control condition that assessed the ability to rapidly make a direct perceptual match (Pattern Matching). In between, participants completed the Adult and Child Emotion recognition conditions, which assessed the ability to recognize facial expressions displayed by adults or by toddlers/infants, respectively. The order of the adult and child conditions was counterbalanced within each group (men, women, parents, non-parents). This was then followed by an Emotion Labelling task, in which participants were asked to verbally identify the emotion shown in each of the infant and toddler images that had been presented earlier, during the Child Emotion condition. The session ended with a test of word meanings (Verbal Meaning Test; Thurstone and Thurstone 1963) and the collection of parenting and other child experience data using a structured questionnaire. Accuracy and response time (RT) were the dependent variables of interest on each of the face discrimination tasks.

Facial Discrimination Tasks

Stimuli for the face discrimination tasks were presented on a desktop PC (3.2 GHz) equipped with 8 GB of RAM (64-bit). Stimulus presentation and recording of response times (RTs) were controlled via E-Prime (E-Prime 2.0, Psychology Software Tools, Sharpsburg, PA). Participants sat 60–75 cm in front of a 60-cm high-resolution LG monitor (1920 × 1080). Images were presented centrally and were a fixed size (22 × 16 cm), except for the Child Emotion condition where images were slightly smaller to better approximate the true size of infant or toddler's faces (about 15 × 14.5 cm). The faces were gray scale digitized photographs from either the Pictures of Facial Affect (Ekman and Friesen 1976) or, for the infant/toddler faces, from the IFEEL Pictures (Infant Facial Expressions of Emotion from Looking at Pictures; Emde et al. 1987) and internet sources. As in Hampson et al. (2006), permuted faces were used in the Pattern Matching condition to provide stimuli that were unrecognizable as faces but preserved all of the contour information available in the original images.

Trials were self-paced. Each trial was initiated by a keypress. Timing of the RT began at image onset and was terminated by a keypress to yield the RT (in milliseconds) on each trial, which was automatically recorded by the computer. Participants were instructed to press the spacebar as quickly as possible while still being accurate. A visual mask was displayed for 20 ms at image offset to terminate further visual processing (Breitneyer 1984). To determine whether a participant had made the correct discrimination on each trial, he/she was instructed to point to the correct choice on a laminated response card placed on an upright stand beside the computer screen. All choices were manually recorded by an examiner. This procedure was used to avoid the inaccuracies in RT measurement commonly introduced on each trial by the use of a six-button horizontal or vertical physical or digital button choice array and the left-to-right or top-to-bottom scanning biases they elicit during decision-making.

The response card for the Adult Emotion Recognition condition is shown in Fig. 1. For each of the four conditions, the response card contained six images (6.75 × 4.5 cm or 6.5 × 5.5 cm) arrayed in two vertical columns of three. Before each condition was administered, participants were shown the response card and were allowed to study the images. Participants signaled when they felt ready to begin.

The following four face discrimination conditions were administered by computer. Within each condition, the order of the trials was randomized individually for each participant. Face Matching was a practice condition only. It was always administered first and allowed participants to practice for speed and to familiarize themselves with the standardized stimulus presentation and response format to be used in each of the emotion recognition conditions that followed. The order of the Adult Emotion and Child Emotion conditions was counterbalanced.

Face Matching

As described in Hampson et al. (2006), this was a practice condition. It required a direct match with the images shown on the response card, but no emotional processing. The faces of six different individuals were presented four times each for a total of 24 trials. All faces had a neutral expression. On each trial, participants pressed the spacebar immediately upon realizing which face had appeared on the screen, then pointed to the matching face on the response card.

Adult Emotion Recognition

This condition was identical to the one used previously by Hampson et al. (2006). Briefly, participants were asked to discriminate emotional expressions depicted on adult faces. On each trial, a face appeared in the center of the screen, displaying one of six basic emotions (Happy, Sad, Fear, Anger, Disgust, Neutral[1]). Neutral faces lacked overt markers of either positive or negative affect (e.g., wrinkling of the nose). There were 60 trials and each image appeared only once. On each trial, the face on the computer was of a different person but displayed the same facial expression as one of the six faces on the response card. The response card showed the six emotions displayed by a single individual who did not appear on any of the 60 trials. Participants were instructed to press the spacebar immediately upon recognizing the emotion that was portrayed, then pointed to the matching emotional expression on the response card. A correct response required participants to recognize the *emotion* depicted. Because a different individual was pictured on the response card, direct perceptual matching of the images based on individual identity was impossible.

Child Emotion Recognition

The Child Emotion Recognition condition was constructed for the present study and paralleled the Adult Emotion Recognition condition in every respect, but employed infants' and toddlers' faces instead of adult faces as stimuli (see Fig. 2 for examples). There were 74 trials, each consisting of a child's face viewed from a frontal position and displaying one of the same six emotions (Happy, Sad, Fear, Anger, Disgust, Neutral). As in the Adult Emotion condition, each face on the computer screen was of a different identity but wore an emotional expression that matched one of the six emotions shown on the response card. The faces on the response card did not appear in any of the 74 trials. Participants pressed the spacebar immediately upon recognizing which emotion was portrayed on each trial, then pointed to the matching expression on the response card.

Pattern Matching

This condition required a direct perceptual match, but non-face stimuli were used. One of the Ekman images was permuted to make the face unrecognizable and two small black rectangles were superimposed on the resulting image. On each

[1] Happiness is the only positive facial expression identified by Ekman and others as innate.

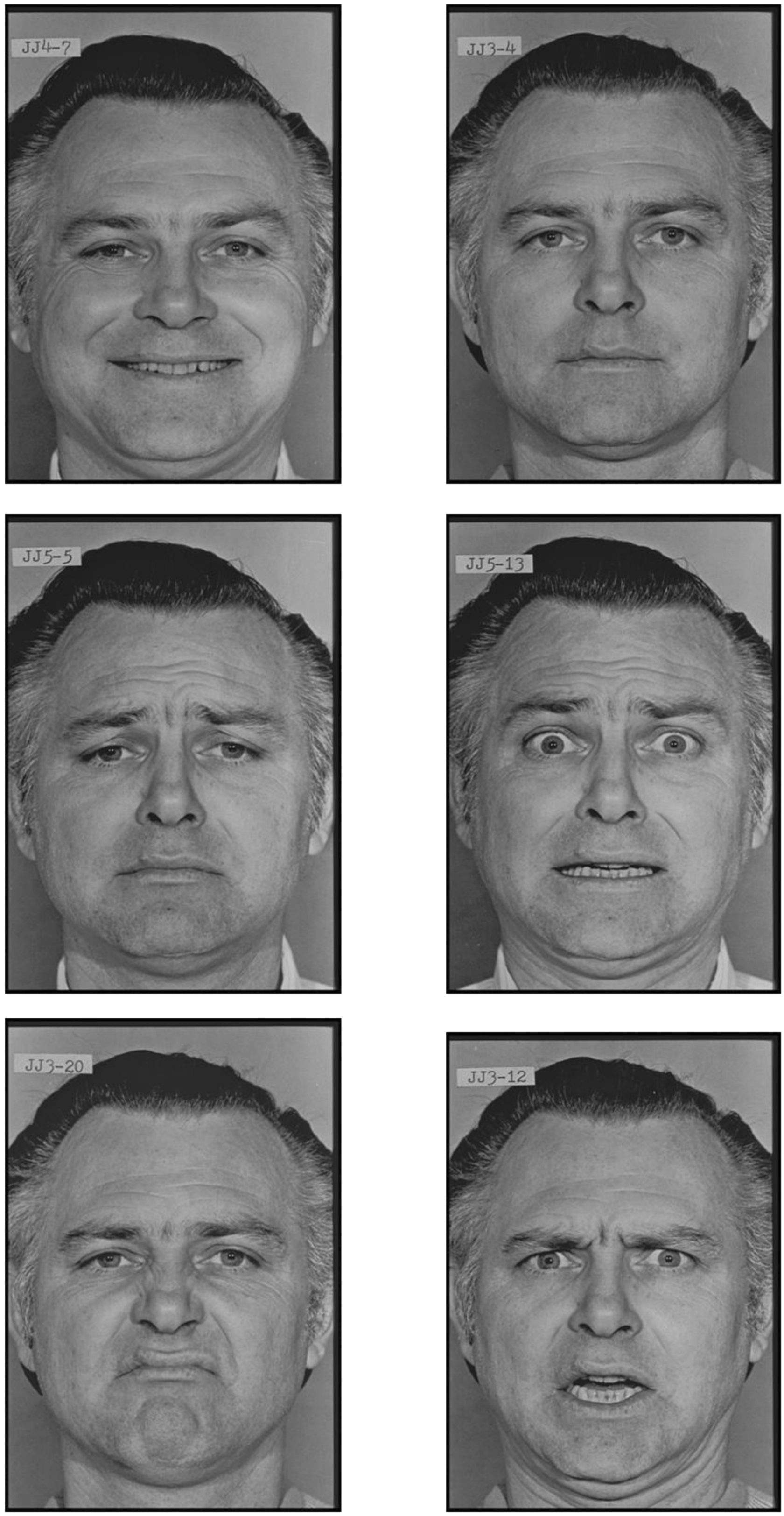


Fig. 1 Example of the response card layout. Shown here is the response card used in the Adult Emotion condition. (Reprinted from *Evolution and Human Behavior*, 27, Hampson E, van Anders SM, Mullin LI, A female advantage in the recognition of emotional facial expressions: test of an evolutionary hypothesis, 401–416 (2006), with permission from Elsevier)

trial, the rectangles appeared at one of six locations upon the image. Participants were asked to press the spacebar as quickly as possible then point to the identical image on the response card. There were 24 trials. This condition served as a control task to establish each individual's baseline level of speed and accuracy of responding.

Fig. 2 Three examples illustrating the type of infant or toddler face images used as stimuli in the Child Emotion condition. A unique image was used on each of the 74 trials and was presented in the center of the computer screen. (Images from iStock.com)

The dependent variables for all of these tasks were RT and accuracy. Accuracy was scored as the percentage (or proportion) of images correctly identified in each condition. RT was scored as the median response time (in milliseconds) based on all trials within each condition where a correct response was made. In the Adult and Child Emotion conditions, separate accuracy and RT scores were calculated for each of the six emotions individually, as well as two composites that reflected performance for the Positive emotions and Negative emotions collectively. One composite (Positive) represented the mean RT (or accuracy) for all conditions in which a positive emotion was presented (Happy or, for the children's faces, Happy and Neutral). Neutral faces were classified as Positive for children's images, because for children, the neutral expressions had a distinctly pleasant valence (Fig. 2). The other composite (Negative) represented the mean RT (or accuracy) averaged across all conditions in which a negative emotion was presented (Fear, Disgust, Anger, Sadness).

Other Tasks

Emotion Labelling

Following the computerized conditions described above, an Emotion Labelling task was given. Each of the emotional images used in the Child Emotion Recognition task was mounted in an album, and participants were asked to verbally name the emotion depicted in each image. Responses were recorded verbatim. Although each image had originally been selected based on the characteristic muscle contraction patterns of the face, the purpose of the Emotion Labelling task was to collect consensus norms as to which of the primary emotions was expressed in each photo. Unlike the procedures often used to derive norms for other image sets (e.g., LoBue and Thrasher 2015; Maack et al. 2017), a forced-choice format (choosing among only the 6 target emotions) was not used, and instead free report was allowed. At the end of the study, the verbal labels supplied by all 95 participants were combined with additional data from another 15 raters not included in the present study, for a total of 110 verbatim labels for each face image. Three independent judges then classified each verbatim response (e.g., "about to cry," "scared"), into one of the six emotional categories (Sad, Fear, Disgust, Anger, Happy, Neutral) or assigned it to a miscellaneous "Other" category if it did not fall within the primary six. Labels were considered semantically equivalent if they were synonymous with the target emotion (e.g., "mad" was considered equivalent to "angry"; "calm" was considered equivalent to "neutral"). Agreement across the three independent judges was high ($M = 96\%$). Based on the resulting norms, 4 images were discarded from consideration (3 blends, 1 duplicate) leaving 70 images that were then scored for RT and accuracy in the Child Emotion Recognition condition described above.

Verbal Meaning Test (Thurstone and Thurstone 1963)

This test was administered to assure there were no chance differences in general ability between the sexes that could affect perceptual scores. Four minutes were allowed to complete 60 items. On each item, the participant had to choose the word from a list of 5 multiple-choice alternatives that matched the meaning of a target word. The score was the number correct (max = 60).

Parenting and Child Experience Questionnaire

Data on parenting and other significant exposure to young children were collected by a structured questionnaire. All participants irrespective of parent status were asked to report the total amount of experience they had with infants and children aged 5 or under in several contexts: (i) taking care of a younger sibling, (ii) other babysitting experience, (iii) taking care of their own child, (iv) as a daycare or preschool worker, (v) as camp counselor or swimming instructor, and (vi) other child experience. For each context, participants rated their cumulative lifetime experience on a scale that ranged from 0 (little or no experience) to 4 (greater than 500 h of cumulative experience), and in addition a total score was derived by summing the ratings across the six contexts.

Any participants who were parents were further asked to indicate how various parenting duties were normally allocated at home between themselves and their partner. Specifically, any participant who reported being a parent was asked to indicate what percentage of the time they (versus their partner) were the one who performed the following basic childcare activities: feeding the child, changing clothing or diapers, bathing the child, playing with child, singing to child, putting child to sleep, carrying child for transport, calming child when angry or upset, and disciplining the child. The percentage reported for each activity was recorded. These nine activities were chosen because they may differentially expose the tending parent to particular emotions on the part of the child and therefore may confer an acquired advantage in recognition through direct experience.

Statistical Analysis

Data for all the Facial Discrimination tasks were screened for outliers prior to statistical analysis, which affected < 1% of the raw data. Wherever possible, mean replacement (Tabachnick and Fidell 1996) was used to rescue any outlying scores identified in order to avoid any reduction in sample size caused by missing datapoints. Each of the three hypotheses we had identified in advance was evaluated independently by a separate ANOVA, to probe for the interaction effects predicted by fitness threat theory (Hampson et al. 2006). Specifically, mixed analysis of variance (ANOVA) with Sex (Male, Female) as a between-subjects factor and Emotional Valence (Positive, Negative) as a within-subjects factor was used to evaluate each hypothesis.[2] To assure the clearest test of the fitness threat theory, only non-parents were included when evaluating Hypotheses 1, 2, and 3. Although conservative, this procedure recognizes that associative learning derived from daily exposure to children confounds any pre-existing adaptations that may be present and complicates interpretation. Thus, all the major analyses included only *non*-parent participants. Age of Face (Adult, Child) was included as an additional, third, factor to test Hypothesis 3. On a purely exploratory basis, we ran one further three-way ANOVA of the children's face data, which incorporated the parent group as well as non-parents and included Parental Status (Parent, Non-parent) as a third factor.

T tests for independent samples were performed to confirm that sex did not influence scores in the control conditions (Face Matching, Pattern Matching) and that the males and females were matched demographically.

Because our hypotheses were theoretically guided and pertained to positive and negative emotions collectively not to specific individual emotions, the composite accuracy or RT scores formed the major dependent variables of interest.

Results

Control Tasks

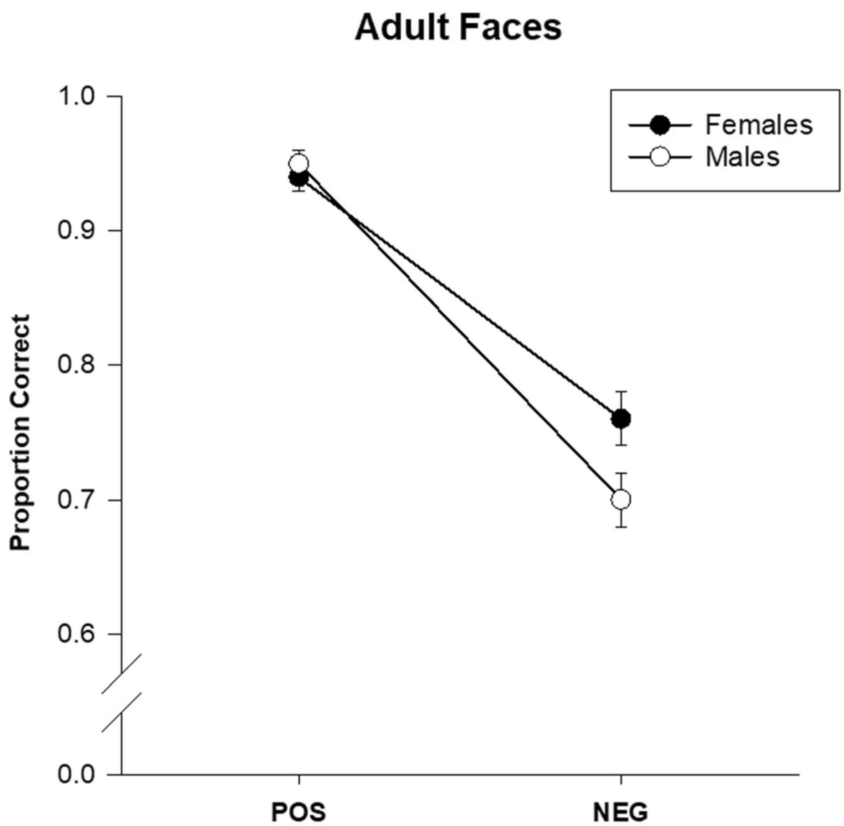
High levels of accuracy were observed in the Face Matching ($M = 96.2\%$, $SD = 5.74$ and $M = 96.5\%$, $SD = 5.66$) and Pattern Matching tasks ($M = 98.8\%$, $SD = 2.29$ and $M = 98.0\%$, $SD = 4.09$) for men and women respectively, indicating a high level of care and attention during the computerized tasks. There was no evidence of a sex difference. Thus, men and women showed highly comparable levels of performance in the two control conditions where no emotional discrimination was required. However, for the two emotional face recognition tasks (described below) sex differences were observed, as predicted. Effects were predominantly evident in accuracy and only rarely in RTs. Accordingly, the description of results for each hypothesis below centers on the accuracy findings, except where RT was relevant.

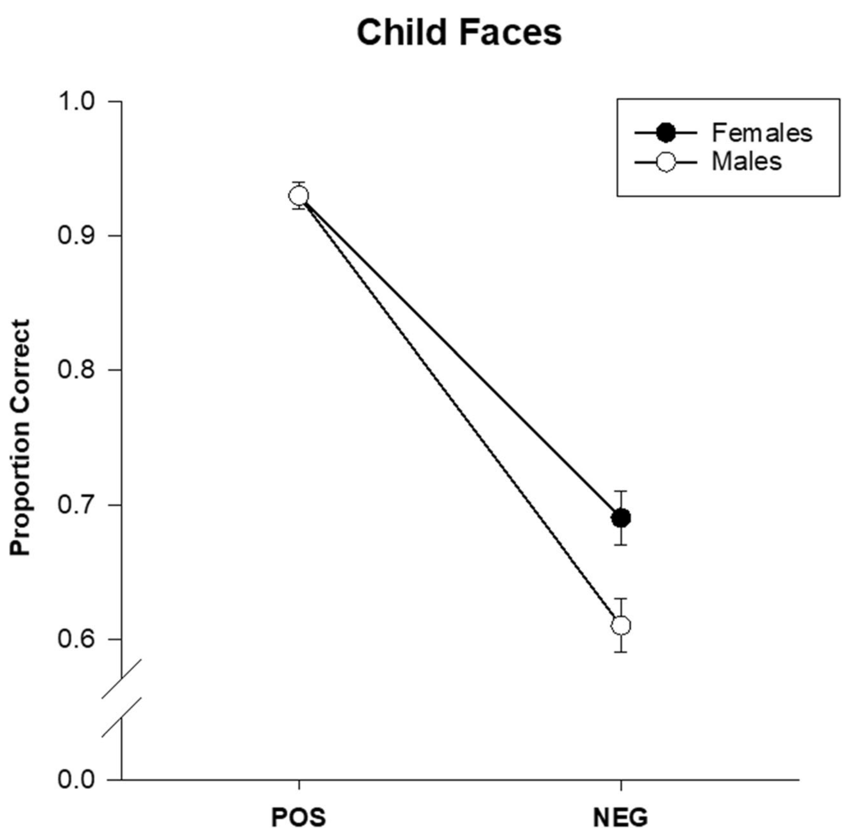
Hypothesis 1: Recognition of Adult Emotional Expressions

To address the first hypothesis, we analyzed whether a sex difference was evident in the ability to discriminate emotion in adult faces. Only non-parents were included in the ANOVA ($n = 36, 35$) to avoid any potential facilitative effects associated with parenting. A mixed factorial ANOVA with Sex (Male, Female) and Emotional Valence as factors revealed a significant main effect of valence. Accuracy was significantly higher for the recognition of positive than negative emotions, F (1, 69) = 284.13, $p < .001$ (Fig. 3, top). More importantly, a significant interaction between sex and valence was found, F (1, 69) = 5.21, $p = .026$ (see Fig. 3). Consistent with the fitness threat hypothesis, women ($M = 0.76$, $SD = 0.09$) showed superior accuracy compared to men ($M = 0.71$, $SD = 0.09$) when identifying the negative expressions that collectively formed the negative composite ($d = 0.56$) ($p < .025$ by a simple effects test, F (1, 69) = 6.67). There was no significant sex difference in the recognition of positive expressions ($M = 0.94$, $SD = 0.07$ and $M = 0.95$, $SD = 0.07$ for women and men, respectively), nor was the overall main effect of sex statistically significant, F (1, 69) = 1.61, $p = .209$.

T tests showed that the male and female groups were closely matched on age ($M = 18.62$, $SD = 1.05$ and $M = 18.64$, $SD = 1.02$, respectively; t (68) = 0.09, $p = .932$) and on general

[2] The ANOVAs were run with one male excluded whose RTs were > 3–5 SDs slower than the mean RT for the sample as a whole. In addition, for the exploratory analysis of the parent data, one female was excluded (a parent) whose response accuracy was below a minimum criterion of 75% correct in the practice condition (Face Matching). Face Matching did not require any emotional decoding whatsoever—the faces presented on the computer merely had to be matched directly by pointing to the same, identical, face image on the response card. Such low accuracy even under direct matching conditions was taken to represent careless or error-prone responding.


Adult Faces
1.0
0.9
0.8
0.7
0.6
0.0
Proportion Correct
POS
NEG
Females
Males


Child Faces
1.0
0.9
0.8
0.7
0.6
0.0
Proportion Correct
POS
NEG
Females
Males

◀ **Fig. 3** Mean proportion of correct discriminations made by females and males in the adult (top) and the infant/toddler (bottom) emotion recognition conditions. A significant sex difference in accuracy was found and was seen only for negatively valenced facial expressions. Bars represent standard error of the mean. POS, positive; NEG, negative

ability as judged by performance on the Verbal Meaning test ($M = 26.14$, $SD = 9.07$ and $M = 23.08$, $SD = 9.27$; t (69) = 1.41, $p = .164$). Demographic differences thus did not explain the sex differences in emotional decoding that were observed.

ANOVA on the median RTs in each of the emotion conditions indicated that positive emotions were correctly identified at shorter latencies than negative ones, F (1, 69) = 40.41, $p < .001$ ($M = 791.89$ ms, $SD = 216.64$ and $M = 983.27$ ms, $SD = 413.94$ for positive and negative emotions respectively). However, the Sex × Valence interaction was not significant for the RTs, F (1, 69) = 0.002, $p = .964$, and there was no main effect of sex, F (1, 69) = 0.77, $p = .382$. Thus, the sex difference in accuracy occurred at comparable response times for the two sexes, and there was no evidence of a speed-accuracy tradeoff in the scores.

Hypothesis 2: Recognition of Children's Emotional Expressions

To evaluate whether there was a sex difference in the recognition of children's facial expressions and whether it conformed to the fitness threat hypothesis, accuracy of decoding also was analyzed for the infant and toddler expressions using a Sex × Valence ANOVA. In order to avoid any effects of learned experience related to parenting, only non-parents were included in the analysis. We had predicted that a female advantage, larger for negative than positive expressions, would be found. Consistent with our a priori prediction, we observed a significant sex by valence interaction when the composite scores were analyzed, F (1, 69) = 5.86, $p = .018$ (Fig. 3, bottom). In agreement with the fitness threat hypothesis, the female group was more accurate ($M = 0.69$, $SD = 0.12$) than the male group ($M = 0.61$, $SD = 0.12$) when identifying negative but not positive emotions ($d = 0.67$). Mean accuracy for each emotion separately is shown in Fig. 4. Although the main effect of Sex was significant also, F (1, 69) = 5.12, $p = .027$, decomposition of the significant Sex × Valence interaction by a simple effects test showed that the sex difference was confined to the negative expressions (i.e., seen only for the negative composite score), F (1, 69) = 13.14, $p < .01$; see Fig. 3). Independent of sex, there was also a main effect of valence, as we had also observed for the adult facial expressions. Mean accuracy was higher for the positive than negative expressions overall, F (1, 69) = 405.09, $p < .001$.

If RT instead of accuracy was used as a dependent variable, there was a main effect of valence (negative expressions were slower to identify), but no main effect of sex, F (1, 69) = 0.24, $p = .629$, nor was the Sex × Valence interaction significant, F (1, 69) = 0.43, $p = .512$.

Hypothesis 3: Magnitude of the Sex Difference

The accuracy findings for children's expressions therefore supported the fitness threat hypothesis. To evaluate Hypothesis 3, a three-way ANOVA (Sex × Valence × Age of Face) was conducted to test the further prediction that the magnitude of the sex effect was enhanced for children's faces, relative to adult faces, because of their greater ecological validity. (Support for Hypothesis 2, confirming the *form* of the core interaction, does not necessarily imply that statistical support for Hypothesis 3 will be found, given that the *magnitude* of the effect is a separate question.) Entering adult and children's faces into a direct comparison showed the same two-way Sex × Valence interaction that was found in the individual ANOVAs described above, F (1, 69) = 9.84, $p = .003$, confirming the relative female advantage in identifying negative facial expressions. However, while the means varied in the expected direction, the three-way interaction was not significant, F (1, 69) = 0.09, $p = .766$. Thus, there was no evidence from the present study that the Sex × Valence interaction was magnified when participants had to identify the facial expressions of young children.

Overall, the adult faces were slightly easier for participants to identify (84% correct for adult faces vs. 79% correct for children's faces), as shown by a significant main effect of Age of Face in the accuracy scores, F (1, 69) = 25.78, $p < .001$.

Exploratory ANOVA: Effects of Parenting

Greater experience with young children acquired through parenting may potentially be associated with greater accuracy in recognizing children's emotional expressions. If parenting does influence facial decoding through learning, then to the extent that differential interaction with children is experienced by each parent in a home, parental experience might disproportionally affect the discriminative accuracy of mothers versus fathers. Therefore, parenthood has a potential to either increase or decrease the magnitude of the female advantage that is seen among young adults who have no parenting experience.

Our sample of self-identified parents was small ($n = 23$). Nevertheless, all study participants with available data ($n = 93$) were entered into an exploratory Parental Status (Parent, Non-parent) × Valence (Positive, Negative) × Sex (Male, Female) ANOVA to test for an effect of parental status on the accuracy scores. Greater experience with children might be expected to affect primarily the decoding of *children's* faces; therefore, only data from the Child Emotion condition were analyzed. The ANOVA confirmed that parents had

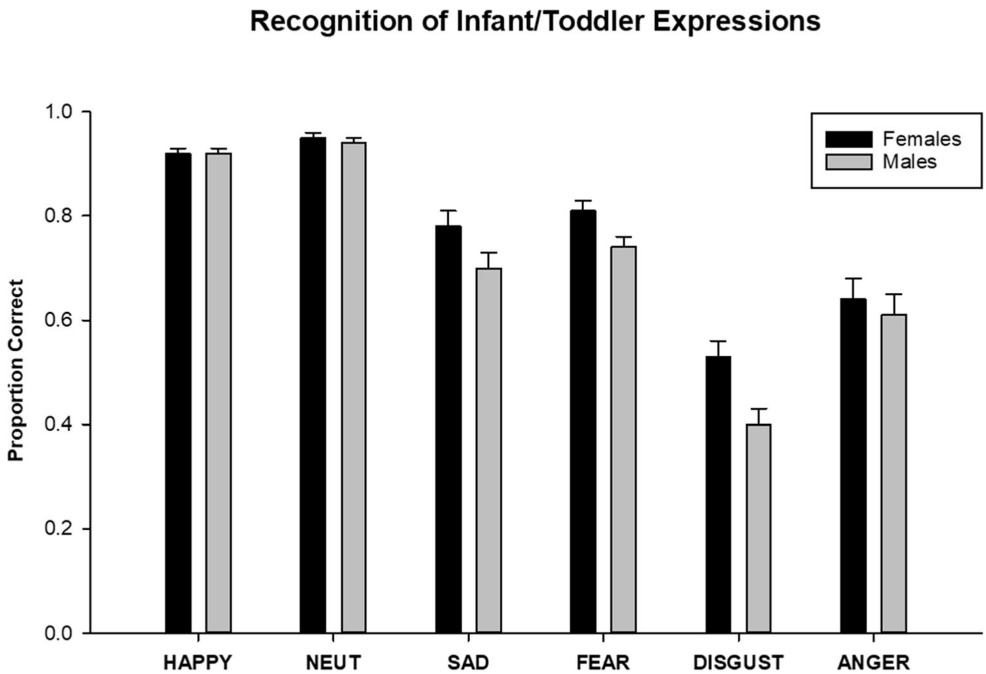


Fig. 4 Mean proportion of correct discriminations made by females and males, shown separately for each of the six categories of emotion in the infant/toddler emotion recognition condition. Bars represent the standard error of the mean

significantly higher accuracy than non-parents when decoding children's facial expressions, F (1, 89) = 5.96, p = .017 (Fig. 5). This was especially true for the negative emotions, Parent × Valence interaction: F (1, 89) = 6.60, p = .012. Specifically, analysis of simple effects indicated that parenthood had no significant influence on the ability to recognize positive expressions, F (1, 89) = 0.07, p = .788 (M = 93–94% correct in all male and female groups regardless of parental status; see Fig. 5). In contrast, for the negative expressions, analysis of simple effects showed that being a parent was associated with significantly higher recognition accuracy, F (1, 89) = 7.55, p = .007, d = 0.72. The improvement in recognition was greatest among males (fathers) (M = 0.71 versus M = 0.61 for fathers and non-fathers respectively, d = 0.77, $p < .01$ by post hoc test; Fig. 5). Among men, the improved accuracy associated with being a father tended to be larger for the negative than positive expressions, Parent × Valence interaction: F (1, 42) = 3.87, p = .056. For women, the same analysis was not statistically significant.

Correlations with Specific Childcare Activities

Spearman's rank order correlation was used to explore potential relationships between exposure to infants and children (aged 5 or under) in various caregiving contexts and the ability to recognize children's facial expressions. We chose the non-parametric Spearman's correlation because several of our experience-related variables were not normally distributed and the scores were ordinal not interval scale. In the sample as a whole (n = 93, which included both the parents and non-parents), being a parent was the only type of acquired experience that significantly predicted the ability to recognize children's facial expressions (see Table 1). The correlation reached significance only for men (r = 0.33, n = 44), but not

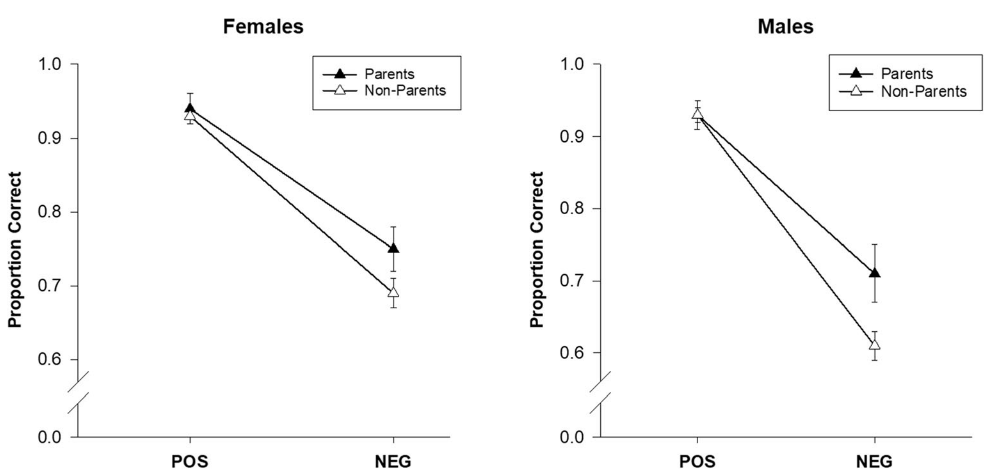


Fig. 5 Mean accuracies of parents and non-parents in the discrimination of infant/toddler emotional expressions. Females are shown on the left, males on the right. Parenting experience was associated with a heightened ability to recognize negative emotions. Bars represent the standard error of the mean. POS, positive; NEG, negative

Table 1 Spearman's correlations between emotional recognition accuracy and the extent of caretaking experience with young children

	Care of sibs	Daycare/preschool worker	Camp counselor or swimming instructor	Raising own child	Babysitting	Other
Males ($n = 44$)						
Positive	.242	.026	.064	.002	.243	− .023
Negative	.114	.045	.076	.333***	.239	.085
Females ($n = 49$)						
Positive	− .135	.028	.099	.022	.040	.050
Negative	.013	.004	− .072	.242	.018	.118

***$p < .05$. Self-reported hours of experience in the various childcare categories shown (e.g., babysitting, care of siblings) ranged from 0 to over 500 h within each category

women ($r = 0.24$, $n = 49$) and only for men's recognition of negative not positive expressions ($r = 0.00$ for the positive expressions; Table 1).

Degree of parental involvement in specific daily childcare activities shared between caregivers (e.g., feeding child, bathing child) could be examined only among the parents ($n = 23$). A few correlations of potential theoretical interest were evident (e.g., parents who took greater responsibility for disciplining their child showed greater accuracy in identifying the emotions anger and disgust), but these correlations were not formally analyzed due to the limited number of parents in the present dataset.

Discussion

A sex difference in the ability to recognize facial expressions of emotion is still controversial even though a recent meta-analysis concluded there is substantial empirical support for this hypothesis (Thompson and Voyer 2014). The present study illustrates why past findings may be inconsistent, in that the magnitude of the sex difference was found to depend on the valence of the emotion displayed. We evaluated the sex difference in healthy men and women of reproductive age, an age when the difference is said to be largest (Thompson and Voyer 2014). Six emotions considered to have universal signal value (Ekman and Friesen 1976), and two control conditions that did not require the discrimination of emotional signals served as stimuli. Consistent with the fitness threat hypothesis, women showed superior accuracy compared to men when discriminating negatively but not positively valenced facial expressions. The sex difference was observed for both adult and infant/toddler faces. The latter may possess greater ecological validity (see below). Direct parental experience was associated with improved accuracy in discriminating infant emotions as reported previously by Proverbio et al. (2007), supporting a potential role for learning or expertise in the decoding of infant facial signals. Importantly, however, the Sex × Valence interaction predicted by the fitness threat hypothesis was observed robustly even in young reproductive age adults who had *no* parental experience. This is consistent with a pre-existing, evolved, adaptation as proposed by the fitness threat hypothesis.

A previous study from our lab (Hampson et al. 2006) was the first to articulate the fitness threat hypothesis and tested it using adult emotional faces as perceptual stimuli. Using various control stimuli, the study demonstrated that a female superiority was not seen for visual processing in general or even for the processing of faces, but was seen only when discrimination of emotion from facial cues was required. Similarly, in the present study, a sex difference was not seen in the control conditions. Babchuk et al. proposed that over the course of hominid evolution female humans evolved specific traits that serve to facilitate their ancestral and near-ubiquitous role of primary caregiver, including heightened facility in identifying facial emotions to promote the survival and care of offspring (Babchuk et al. 1985). Data consistent with the caretaker hypothesis have been reported using several experimental paradigms (e.g. Hampson et al. 2006; Williams et al. 2009; Wingenbach et al. 2018). However, Hampson et al. (2006) found that the sex difference was magnified for negatively valenced facial expressions and emphasized the importance of fitness threat in the evolution of the difference. Specifically, women tested by Hampson et al. (2006) showed greater facility than men when decoding negative emotions, resulting in a statistical interaction between sex and emotional valence. This interaction implies that negative expressions have been subject to differential selection pressures. A dissociation between different emotions is plausible in light of recent neurobiological studies showing that the cerebral representation of emotion in the human brain is highly complex, with activation of both common and specific brain sites in response to different emotions (Adolphs et al. 1996; Phan et al. 2002; Weisenbach et al. 2014; Willinger et al. 2019).

Negative expressions in the present study were generally harder to discriminate than positive ones, a pattern found in both sexes in this and previous studies (Ekman and Friesen 1976; Kuharic et al. 2019; Thompson and Voyer 2014; Williams et al. 2009). Nonetheless, the current study replicated the interaction previously observed by Hampson et al. (2006) in a new sample of young adults. Women were significantly more accurate than men at identifying negative expressions (and/or quicker to identify the correct emotion, Hampson et al. 2006) as predicted by the fitness threat hypothesis. In both studies, the decisions were self-paced. Participants were instructed to respond as quickly as possible while still being accurate. Despite this instruction, in our 2006 study, the mean RT in each condition was notably slower than in the present work (up to 1200 ms in Fig. 2 of Hampson et al. 2006, versus 800–900 ms in the present study). A difference in average speed of response might explain why a ceiling effect in accuracy was evident in our 2006 work, but why accuracy was able to reveal the interaction effect here. A tradeoff between RT and accuracy is commonly seen in reaction time studies, and the RT is a superior measure in situations where accuracy approaches ceiling (Coren et al. 1994). In the present work, the predicted interaction between sex and valence was successfully demonstrated using accuracy as a dependent variable.

The fitness threat hypothesis argues that greater female ease in identifying emotional expressions stems from greater parental investment by females (Trivers 1972) combined with a need for the primary caretaker to accurately decode emotional signals displayed by preverbal offspring that signal a need for action. It is important, therefore, to test the fitness threat hypothesis using the most ecologically valid perceptual stimuli, i.e., emotional expressions displayed by infants and toddlers. To our knowledge, the present study is the first to test the fitness threat hypothesis by using infants' and toddlers' expressions.

Children's faces are often found to be more difficult to decode than adult faces, with some studies finding that both adults and adolescents are more accurate at identifying the emotions displayed by adult faces (Marusak et al. 2013; Thompson and Voyer 2014). In the present study, this was evident as a significant main effect, whereby slightly higher accuracy was found for adult expressions. Nonetheless, the predicted interaction between sex and valence was also observed. Just as we saw for adult faces, women displayed higher accuracy than men for decoding negative but not positive expressions made by young children. This was evident when the positive and negative composite scores were statistically compared but also for most of the negative emotions considered individually (Fig. 4). In contrast, no sex difference was seen under control conditions (Face Matching, Pattern Matching) that did not require emotional decoding. Thus, children's faces, too, robustly supported the fitness threat hypothesis. In fact, there was no evidence in the present study of any sex difference in the decoding of positive expressions. This is contrary to what might be predicted based on other, laxer, interpretations of the primary caretaker hypothesis, including the idea that the sex difference evolved to promote attachment relationships (see Hampson et al. 2006, for a detailed discussion).

Although the sex by valence interaction predicted by the fitness threat hypothesis was clear and apparent for children's faces, we did not find support for our further prediction that the magnitude of the interaction may be accentuated under conditions where children's faces must be decoded. Although the size of the sex disparity for negative expressions was slightly larger in absolute terms if children's rather than adults' faces were presented, this effect was not significant (as evidenced by the absence of a three-way interaction between sex, valence, and age of the face) and was very modest in size (about 3% larger for infant/toddler stimuli). The greater difficulty observed in this study and many others when decoding children's faces might mitigate against finding an accentuated sex difference. Alternatively, it is possible that whatever perceptual mechanisms have evolved respond equally well to the distinctive facial configurations that characterize the six primary emotions irrespective of the age of the facial background on which an emotion is expressed. Indeed, recent work found that the ability to identify specific emotions (e.g., fear) in adult faces correlated with judgments of emotion in similarly valenced infant faces (Parsons et al. 2019), suggesting there might exist a common expression-specific emotion recognition mechanism across both infant and adult faces. Future work might further address this theoretical question, perhaps via a different experimental paradigm that is able to more perfectly equate the perceived difficulty level of the adult and child face stimuli.

The current study is the first to test the fitness threat hypothesis by using children's expressions. Our basic findings, however, are broadly consistent with general trends seen in studies of adult faces (e.g., Anderson et al. 2011; Williams et al. 2009; for review, see Thompson and Voyer 2014) and recent work by Proverbio et al. (2007) that involved infant expressions, who noted that women were more accurate than men when responding to expressions that were "weakly negative" (see also Hoffmann et al. 2010). An earlier study by Babchuk et al. (1985) found an across-the-board female advantage in the recognition of children's facial expressions but did not test for any differences between expressions having a positive or negative valence. In the present study, a valence-dependent sex difference was found for both posed expressions (Pictures of Facial Affect, Ekman and Friesen 1976) and spontaneous ones (infant and toddler faces), suggesting that the spontaneity dimension is not material to seeing the valence effect. This too is consistent with past work showing that the use of natural versus posed expressions ("intention of actors") does not affect the decoding of adult faces (Thompson and Voyer 2014).

From a skeptical point of view, one might ask whether the valence-dependent sex difference we observed could be just an artifact of differences in the ease of identification of positive versus negative expressions and thus not meaningful at a theoretical level. This seems quite unlikely. First, our earlier study (Hampson et al. 2006) documented exactly the same sex by valence interaction as in the present work, but used RT rather than accuracy as the primary dependent variable. As such, the sex difference found by Hampson et al. (2006) is not readily explained in terms of ceiling or floor effects. The exact same interaction was evident in the present study, albeit expressed in terms of accuracy. Secondly, in Hampson et al. (2006), accuracy scores calculated for disgust and anger were statistically analyzed and showed the same female advantage seen here, as reported briefly by the authors for those two emotions (Hampson et al. 2006, p. 408). The female advantage was seen even though the accuracies for disgust and anger were as high as 85–90% (Hampson et al. 2006), a percentage comparable to accuracy for happy infant stimuli in the present study (cf. Fig. 4, present study). Thus, stimuli closely matched on accuracy level showed *no* sex difference in the current study (happy) but *did* show a female advantage, in the predicted direction, of approximately 0.5 *SD* in the previous study (anger, disgust). This suggests that the sex difference is, in fact, valence-dependent. Thirdly, it can be pointed out that men in the present study had plenty of scope to do more *poorly* than women when judging the positive expressions, yet they did not. Instead, a sex difference emerged selectively for the negatively valenced expressions, where overall accuracies were in the middle range. This suggests there is something special about the emotional valence that results in the expression of the sex effect.

Are there alternative theories that could explain the valence-dependent female advantage? Adult men and women differ in average body size and engagement in physical aggression (Sell et al. 2012), but avoidance of male violence is unlikely to be a basis for the perceptual advantage we observed, because it predicts a female advantage only for the detection of anger. Other theories invoking harm-avoidance to explain female adaptations or facial preferences (e.g., Sacco et al. 2017) do not readily apply to negative emotions like sadness, nor explain why a heightened capacity to identify negative emotions would be seen so robustly for infant and toddlers' faces, who pose no threat of harm (and in fact are vulnerable to harm themselves). Conversely, if we suppose that selection has operated on males instead of females, it is not obvious what adaptive purpose could be served by a *blunted* ability to identify negative emotions in men. A suite of adaptations to promote male combat-readiness has been proposed by theories emphasizing the ancestral importance of inter-male physical conflict (Sell et al. 2012). Even if we extend this idea to encompass reduced emotional reactivity, a *male* superiority in the capacity to recognize fear and anger might still be predicted to exist (and in fact has been argued by some authors, e.g., Rotter and Rotter 1988) because sensitivity to these important social signals can offer a competitive advantage in reading one's opponents. Empirically, however, it was a female not male superiority that was observed in the present study. Therefore, our data seem best conceptualized in terms of an adaptation to promote effective child-rearing.

A final, exploratory, analysis in the present study pertained to expertise effects and their potential implications for finding a sex difference. We tested if greater experience with young children acquired through parenting would be associated with increased accuracy in the decoding of children's expressions. Parents in the present study did in fact exhibit greater accuracy than non-parents, supporting the idea that parenthood confers an experience- or expertise-dependent effect. Importantly, though, while experience with children did affect the magnitude of the sex difference (at least for negative emotions), limiting our analysis to non-parents showed that the sex difference was robustly apparent even in young reproductive age adults who had no parenting experience whatsoever.

Although our sample size was small, it was fathers who benefited most from parenting practice. This will need to be confirmed in an independent sample in a future study using a larger sample size. At least in the present study, fatherhood was associated with a sharpened ability to accurately recognize negative emotions compared with other males (Fig. 5). Women on the other hand were not significantly affected by parenting experience (there was at best a very weak statistical trend). If confirmed by future work, the fact that women were less affected than men by parenting is intriguing because it is consistent with the idea that the superior accuracy they enjoy derives from an evolved adaptation that affords women superior recognition accuracy even in the absence of acquired parenting experience. How such an adaptation comes about, and whether or not other life experiences (including those occurring early in childhood; McClure 2000) also play a role in bringing the sex difference to full expression, is currently unknown. In other species, many sex differences originate as a direct result of exposure of the nervous system to key hormones before birth (see Breedlove and Hampson 2002, for a review) and are further elaborated at puberty.

In the present study, only parenting, not other forms of direct child experience, was significantly associated with the ability to recognize infant/toddlers' facial expressions (Table 1). However, smaller correlations for other categories of interactive child experience were also observed (e.g., babysitting). The present data do not exclude the possibility that other forms of childcare experience could contribute to increased facial decoding competency. They do, however, point to the possible importance of parenting and the intensive practice it provides. In our earlier study (Hampson et al. 2006), we did not find any associations between the female advantage and the amount of previous childcare experience that our participants reported, but in that study we did not assess parenting, and the decoding of infant/toddler faces could not be assessed because only adult face stimuli were used (see also Babchuk et al. 1985).

We note that an effect of parenthood need not always be based on learning. In other species and potentially also in humans, pregnancy and motherhood are associated with significant biological changes in females, some of them initiated by ante- or postpartum endocrine events that can trigger or enhance behavioral, perceptual, and cognitive adaptations that equip females for motherhood by activating brain-related maternal adaptations designed to promote infant survival and success (e.g., Macbeth and Luine 2010; Ouellette and Hampson 2019). Neurobiological changes in fathers also occur, ranging from innate adaptations to fatherhood (Feldman et al. 2019) to changes thought to be triggered by responsive caregiving through a bottom-up process (Abraham et al. 2014). Thus, improved emotional processing associated with being a parent can, but need not, signify a learning or experiential mechanism. Indeed, differences between the brains of parents and non-parents have been documented by recent neuroimaging or electrophysiological studies (Abraham et al. 2014; Peltola et al. 2014; Proverbio et al. 2006; Seifritz et al. 2003).

The present study adds to the literature on facial affect recognition by establishing that there is a larger female superiority for the decoding of negative than positive facial expressions and that it precedes and is not a consequence of learned experience associated with motherhood. Future studies may gain new functional insights into mechanisms by employing signal detection analysis (Stanislaw and Todorov 1999) to measure sensitivity (d') and/or examine criterion differences that might exist between men and women. Our findings help to explain some of the inconsistencies of past literature, where a female superiority in emotional discrimination is not found universally. The sex difference and its magnitude may depend in part on the affective valence of the specific emotions that are assessed. The current data provide new and important support for the fitness threat hypothesis (Hampson et al. 2006), which argues that a female processing advantage, heightened for the perception of negative expressions, is an evolved adaptation related to greater ancestral female investment in parental care and the tending of preverbal offspring. The heightened ability to accurately identify negative emotions may be a mechanism to enhance reproductive success in women.

Availability of Data and Materials The dataset analyzed and reported in the current study is available upon request from the corresponding author.

Author Contributions EH was responsible for study conception, study design, and data interpretation. Material preparation and writing of E-Prime scripts was done by EH, SO, or former research assistants in the laboratory. SO, JAC, SJO, PI performed data collection and scoring. PI and BH assisted with norms development for the verbal labelling task. Statistical analysis was performed by EH, JAC, SO. The first draft of the paper was written by EH. All authors commented on previous versions of the manuscript and have read and approved the final version. EH was responsible for funding acquisition and resources, and supervised the data collection and analysis.

Funding Information S. Owais is now a MD/PhD student at McMaster University, B. Howidi is a graduate student at McGill, and J. Chow is a medical student at the University of Alberta. This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Grant No. 138017 to E. Hampson. P. Istasy was supported by an NSERC Undergraduate Student Research Award.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflicts of interest.

Ethics Approval This study received institutional ethics approval from the University of Western Ontario Non-Medical Research Ethics Board and was carried out in compliance with the Declaration of Helsinki.

Consent to Participate All participants gave written informed consent to participate.

References

Abraham, E., Hendler, T., Shapira-Lichter, I., Kanat-Maymon, Y., Zagoory-Sharon, O., & Feldman, R. (2014). Father's brain is sensitive to childcare experiences. *Proceedings of the National Academy of Sciences, 111*, 9792–9797.

Adolphs, R., Damasio, H., Tranel, D., & Damasio, A. R. (1996). Cortical systems for the recognition of emotion in facial expressions. *Journal of Neuroscience, 16*, 7678–7687.

Anderson, I. M., Shippen, C., Juhasz, G., Chase, D., Thomas, E., Downey, D., Toth, Z. G., Lloyd-Williams, K., Elliott, R., & Deakin, J. F. W. (2011). State-dependent alteration in face emotion recognition in depression. *British Journal of Psychiatry, 198*, 302–308.

Babchuk, W. A., Hames, R. B., & Thompson, R. A. (1985). Sex differences in the recognition of infant facial expressions of emotion: the primary caretaker hypothesis. *Ethology and Sociobiology, 6*, 89–101.

Breedlove, S. M., & Hampson, E. (2002). Sexual differentiation of the brain and behavior. In J. B. Becker, S. M. Breedlove, D. Crews, & M. M. McCarthy (Eds.), *Behavioral endocrinology* (pp. 75–114). Cambridge: MIT Press.

Breitneyer, B. G. (1984). *Visual masking: an integrative approach*. Oxford: Clarendon Press.

Calder, A. J., Keane, J., Manly, T., Sprengelmeyer, R., Scott, S., Nimmo-Smith, I., & Young, A. W. (2003). Facial expression recognition across the adult life span. *Neuropsychologia, 41*, 195–202.

Connolly, H. L., Lefevre, C. E., Young, A. W., & Lewis, G. J. (2019). Sex differences in emotion recognition: evidence for a small overall female superiority on facial disgust. *Emotion, 19*, 455–464.

Coren, S., Ward, L. M., & Enns, J. T. (1994). *Sensation and perception* (4th ed.). Toronto: Harcourt Brace.
Ekman, P. (1994). Strong evidence for universals in facial expressions: a reply to Russell's mistaken critique. *Psychological Bulletin, 115*, 268–287.
Ekman, P. (1997). Should we call it expression or communication? *Innovation, 10*, 333–344.
Ekman, P., & Friesen, W. V. (1976). *Pictures of facial affect*. San Francisco, CA: Human Interaction Laboratory, University of California Medical Center.
Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., et al. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology, 53*, 712–717.
Emde, R. N., Butterfield, P. M., & Osofsky, J. D. (1987). *I feel pictures: infant facial expressions of emotion from looking at pictures*. Denver: University of Colorado.
Feldman, R., Braun, K., & Champagne, F. A. (2019). The neural mechanisms and consequences of paternal caregiving. *Nature Reviews Neuroscience, 20*, 205–224.
Grimshaw, G. M., Bulman-Fleming, M. B., & Ngo, C. (2004). A signal-detection analysis of sex differences in the perception of emotional faces. *Brain and Cognition, 54*, 248–250.
Hampson, E., van Anders, S. M., & Mullin, L. I. (2006). A female advantage in the recognition of emotional facial expressions: test of an evolutionary hypothesis. *Evolution and Human Behavior, 27*, 401–416.
Hoffmann, H., Kessler, H., Eppel, T., Rukavina, S., & Traue, H. C. (2010). Expression intensity, gender and facial emotion recognition: women recognize only subtle facial emotions better than men. *Acta Psychologica, 135*, 2778–2283.
Izard, C. E. (1994). Innate and universal facial expressions: evidence from developmental and cross-cultural research. *Psychological Bulletin, 115*, 288–299.
Javanbakht, A., King, A. P., Evans, G. W., Swain, J. E., Angstadt, M., Phan, K. L., & Liberzon, I. (2015). Childhood poverty predicts adult amygdala and frontal activity and connectivity in response to emotional faces. *Frontiers in Behavioral Neuroscience, 9*, 154. https://doi.org/10.3389/fnbeh.2015.00154.
Kuharic, D. B., Makaric, P., Kekin, I., Lovrencic, I. L., Savic, A., Ostojic, D., et al. (2019). Differences in facial emotional recognition between patients with the first-episode psychosis, multi-episode schizophrenia, and healthy controls. *Journal of the International Neuropsychological Society, 25*, 165–173.
LoBue, V., & Thrasher, C. (2015). The Child Affective Facial Expression (CAFE) set: validity and reliability from untrained adults. *Frontiers in Psychology, 5*, 1532. https://doi.org/10.3389/fpsyg.2014.01532.
Lopata, C., Thomeer, M. L., Volker, M. A., Lee, G. K., Smith, T. H., Smith, R. A., Mcdonald, C. A., Rodgers, J. D., Lipinski, A. M., & Toomey, J. A. (2012). Feasibility and initial efficacy of a comprehensive school-based intervention for high-functioning autism spectrum disorders. *Psychology in the Schools, 49*, 963–974.
Maack, J. K., Bohne, A., Nordahl, D., Livsdatter, L., Lindahl, A. A. W., Øvervoll, M., Wang, C. E. A., & Pfuhl, G. (2017). The Tromsø Infant Faces Database (TIF): development, validation and application to assess parenting experience on clarity and intensity ratings. *Frontiers in Psychology, 8*, 409. https://doi.org/10.3389/fpsyg.2017.00409.
Macbeth, A. H., & Luine, V. N. (2010). Changes in anxiety and cognition due to reproductive experience: a review of data from rodent and human mothers. *Neuroscience and Biobehavioral Reviews, 34*, 452–467.
Mandal, M. K., & Palchoudhury, S. (1985). Perceptual skill in decoding facial affect. *Perceptual and Motor Skills, 60*, 96–98.
Marusak, H. A., Carré, J. M., & Thomason, M. E. (2013). The stimuli drive the response: an fMRI study of youth processing adult or child emotional face stimuli. *NeuroImage, 83*, 679–689.
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.
McNair, D. M., Lorr, M., & Droppleman, L. F. (1971). *Profile of mood states*. San Diego: EdITS.
Nishitani, S., Doi, H., Koyama, A., & Shinohara, K. (2011). Differential prefrontal response to infant facial emotions in mothers compared with non-mothers. *Neuroscience Research, 70*, 183–188.
Ouellette, S. J., & Hampson, E. (2019). Memory and affective changes during the antepartum: a narrative review and integrative hypothesis. *Journal of Clinical and Experimental Neuropsychology, 41*, 87–107.
Parsons, C.E., Nummenmaa, L., Sinerva, E., Korja, R., Kajanoja, J., Young, K.S., Karlsson H., Karlsson L. (2019). Investigating the effects of perinatal status and gender on adults' responses to infant and adult facial emotion. *Emotion*. https://doi.org/10.1037/emo0000698.
Peltola, M. J., Yrttiaho, S., Puura, K., Proverbio, A. M., Mononen, N., Lehtimäki, T., & Leppänen, J. M. (2014). Motherhood and oxytocin receptor genetic variation are associated with selective changes in electrocortical responses to infant facial expressions. *Emotion, 14*, 469–477.
Phan, K. L., Wager, T., Taylor, S. F., & Liberzon, I. (2002). Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI. *NeuroImage, 16*, 331–348.
Proverbio, A. M., Brignone, V., Matarazzo, S., Del Zotto, M., & Zani, A. (2006). Gender and parental status affect the visual cortical response to infant facial expression. *Neuropsychologia, 44*, 2987–2999.
Proverbio, A. M., Matarazzo, S., Brignone, V., Del Zotto, M., & Zani, A. (2007). Processing valence and intensity of infant expressions: the roles of expertise and gender. *Scandinavian Journal of Psychology, 48*, 477–485.
Rotter, N. G., & Rotter, G. S. (1988). Sex differences in the encoding and decoding of negative facial emotions. *Journal of Nonverbal Behavior, 12*, 139–148.
Russell, J. A., Bachorowski, J., & Fernández-Dols, J. M. (2003). Facial and vocal expressions of emotion. *Annual Review of Psychology, 54*, 329–349.
Russo, M., Mahon, K., Shanahan, M., Solon, C., Ramjas, E., Turpin, J., & Burdick, K. E. (2015). The association between childhood trauma and facial emotion recognition in adults with bipolar disorder. *Psychiatry Research, 229*, 771–776.
Sacco, D. F., Brown, M., Lustgraaf, C. J. N., & Young, S. G. (2017). Women's dangerous world beliefs predict more accurate discrimination of affiliative facial cues. *Evolutionary Behavioral Sciences, 11*, 309–315.
Seifritz, E., Esposito, F., Neuhoff, J. G., Lüthi, A., Mustovic, H., Dammann, G., von Bardeleben, U., Radue, E. W., Cirillo, S., Tedeschi, G., & di Salle, F. (2003). Differential sex-independent amygdala response to infant crying and laughing in parents versus nonparents. *Biological Psychiatry, 54*, 1367–1375.
Sell, A., Hone, L. S. E., & Pound, N. (2012). The importance of physical strength to human males. *Human Nature, 23*, 30–44.
Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*, 137–149.
Székely, E., Lucassen, N., Tiemeier, H., Bakermans-Kranenburg, M. J., Van Ijzendoorn, M. H., Kok, R., et al. (2014). Maternal depressive symptoms and sensitivity are related to young children's facial expression recognition: the Generation R Study. *Development and Psychopathology, 26*, 333–345.
Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.). New York: HarperCollins.
Thompson, A. E., & Voyer, D. (2014). Sex differences in the ability to recognize non-verbal displays of emotion: a meta-analysis. *Cognition and Emotion, 28*, 1164–1195.

Thompson-Booth, C., Viding, E., Mayes, L. C., Rutherford, H. J. V., Hodsoll, S., & McCrory, E. J. (2014). Here's looking at you, kid: attention to infant emotional faces in mothers and non-mothers. *Developmental Science, 17*(1), 35–46.

Thurstone, L. L., & Thurstone, T. G. (1963). *Primary mental abilities*. Chicago: Science Research Associates.

Tottenham, N. (2013). Early social deprivation and the neurobiology of interpreting facial expressions. In M. R. Banaji & S. A. Gelman (Eds.), *Navigating the social world: what infants, children, and other species can teach us* (pp. 155–160). New York: Oxford University Press.

Trivers, R. (1972). Parental investment and sexual selection. In B. Campbell (Ed.), *Sexual selection and the descent of man* (pp. 136–179). Chicago: Aldine.

Van Hooff, J. A. R. A. M. (1976). The comparison of facial expressions in man and higher primates. In M. Van Cranach (Ed.), *Methods of interference from animal to human behavior* (pp. 165–196). Chicago: Aldine.

Van Rheenen, T. E., & Rossell, S. L. (2014). Let's face it: facial emotion processing is impaired in bipolar disorder. *Journal of the International Neuropsychological Society, 20*, 200–208.

Weisenbach, S. L., Rapport, L. J., Briceno, E. M., Haase, B. D., Vederman, A. C., Bieliauskas, L. A., Welsh, R. C., Starkman, M. N., McInnis, M. G., Zubieta, J. K., & Langenecker, S. A. (2014). Reduced emotion processing efficiency in healthy males relative to females. *Social Cognitive and Affective Neuroscience, 9*, 316–325.

Williams, L. M., Mathersul, D., Palmer, D. M., Gur, R. C., Gur, R. E., & Gordon, E. (2009). Explicit identification and implicit recognition of facial emotions: I. Age effects in males and females across 10 decades. *Journal of Clinical and Experimental Neuropsychology, 31*, 257–277.

Willinger, D., Karipidis, I. I., Beltrani, S., DiPietro, S. V., Sladky, R., Walitza, S., et al. (2019). Valence-dependent coupling of prefrontal-amygdala effective connectivity during facial affect processing. *eNeuro, 6*, ENEURO.0079–ENEU19.2019. https://doi.org/10.1523/ENEURO.0079-19.2019.

Wingenbach, T. S. H., Ashwin, C., & Brosnan, M. (2018). Sex differences in facial emotion recognition across varying expression intensity levels from videos. *PLoS One, 13*, e0190634. https://doi.org/10.1371/journal.pone.0190634.