Developmental Science

Developmental Science 18:6 (2015), pp 926–939 DOI: 10.1111/desc.12281

PAPER

Mapping the development of facial expression recognition

Helen Rodger, Luca Vizioli, Xinyi Ouyang and Roberto Caldara*

Department of Psychology, University of Fribourg, Switzerland

Abstract

Reading the non-verbal cues from faces to infer the emotional states of others is central to our daily social interactions from very early in life. Despite the relatively well-documented ontogeny of facial expression recognition in infancy, our understanding of the development of this critical social skill throughout childhood into adulthood remains limited. To this end, using a psychophysical approach we implemented the QUEST threshold-seeking algorithm to parametrically manipulate the quantity of signals available in faces normalized for contrast and luminance displaying the six emotional expressions, plus neutral. We thus determined observers' perceptual thresholds for effective discrimination of each emotional expression from 5 years of age up to adulthood. Consistent with previous studies, happiness was most easily recognized with minimum signals (35% on average), whereas fear required the maximum signals (97% on average) across groups. Overall, recognition improved with age for all expressions except happiness and fear, for which all age groups including the youngest remained within the adult range. Uniquely, our findings characterize the recognition trajectories of the six basic emotions into three distinct groupings: expressions that show a steep improvement with age – disgust, neutral, and anger; expressions that show a more gradual improvement with age – sadness, surprise; and those that remain stable from early childhood – happiness and fear, indicating that the coding for these expressions is already mature by 5 years of age. Altogether, our data provide for the first time a fine-grained mapping of the development of facial expression recognition. This approach significantly increases our understanding of the decoding of emotions across development and offers a novel tool to measure impairments for specific facial expressions in developmental clinical populations.

Research highlights

- Our data provide a fine-grained mapping of the development of facial expression recognition for all six basic emotions and a neutral expression.
- Model fitting revealed that the developmental trajectories of facial expression recognition followed three trends: Disgust, neutral and anger expressions showed a steep improvement across development; sadness and surprise showed a more gradual improvement; whereas recognition of happiness and fear remained stable from early childhood, suggesting that the coding for these expressions is already mature by 5 years of age.
- Two main phases were identified in the development of facial expression recognition, ranging from 5 to 12 years old and 13 years old to adulthood.
- This approach offers a novel psychophysical tool to measure impairments for specific facial expressions in developmental clinical populations.

Introduction

The ability to accurately decode complex emotional cues in our social environment is a defining feature of human cognition and is essential for normative social development. How we recognize and process facial expressions of emotion throughout development to reach maturity in adulthood is a pivotal question for developmental psychologists, neuroscientists, educators, and caregivers alike, who aim to trace both typical and atypical trajectories of this important social skill. Despite the relatively well-documented developmental course of emotion recognition in infancy, it is acknowledged that our understanding of the development of this important social function throughout childhood, particularly after the preschool years, remains limited (Mancini, Agnoli, Baldaro, Bitti & Surcinelli, 2013; Thomas, De Bellis, Graham & LaBar, 2007). This enduring gap in the literature is surprising, especially for this stage of development as opportunities for social learning increase

Address for correspondence: Roberto Caldara, Department of Psychology, University of Fribourg, R. Faucigny 2, 1700 Fribourg, Switzerland; e-mail: roberto.caldara@unifr.ch

greatly with the onset of school (Johnston, Kaufman, Bajic, Sercombe, Michie *et al.*, 2011) and evidence suggests that the ability to recognize facial expressions at age 5 predicts later social and academic competence (Izard, Fine, Schultz, Mostow, Ackerman *et al.*, 2001). In one of the most recent reviews of the development of facial expression recognition (FER) during childhood and adolescence, Herba and Phillips (2004) specify the need for normative data across this age range, not only for a greater understanding of this vital social function throughout development, but also to aid identification of atypical emotional development. They further stress the need for studies examining the continued development of emotional expression recognition from childhood through adolescence into early adulthood, as very little is known about development across the full childhood range up to adulthood.

Infant and childhood behavioral studies of facial expression recognition have naturally employed diverse methods according to the presence or absence of language, making comparisons across these groups difficult. Two common approaches employed during the stage of interest here, from early childhood onwards, attempt to minimize language ability confounds by using minimal verbal communication. Matching and labeling tasks require participants either to match emotional expressions from an array of expressions, or to label emotional expressions in a forced choice paradigm or freely without constraints (Mondloch, Geldart, Maurer & Le Grand, 2003). Across such studies there is general agreement that happiness is most accurately recognized at the youngest age, while fear is consistently one of the most difficult expressions to recognize. There is less agreement concerning the trajectory of the other basic emotions following mixed reports in the literature; sadness and anger are frequently cited as being recognized most accurately and earliest subsequent to happiness, followed by surprise and disgust (Herba & Phillips, 2004; Widen, 2013). Some of the discrepancies reported for developmental rates of expression recognition can be accounted for by task effects, as FER performance has been shown to be task dependent (Vicari, Snitzer Reilly, Pasqualetti, Vizzotto & Caltagirone, 2000; Montirosso, Peverelli, Frigerio, Crespi & Borgatti, 2010; Johnston *et al.*, 2011). For example, even within the same study very different results can be found for the same expression depending on the task employed (Vicari *et al.*, 2000). In this instance, performance for disgust across all age groups in the labeling task was much lower than in the matching task. This difference is most likely attributed to differing language demands of the labeling versus matching task, particularly as this trend was beginning to narrow in the oldest age group (9 to 10 years) possessing greater language ability.

A third behavioral approach has more recently been employed to examine children's facial expression recognition accuracy as a function of expression intensity. The motivation for such studies is derived from the fact that we frequently see more subdued expressions of emotion in daily life; therefore do older children recognize subtle expressions of emotion more easily? Again, the results reported in the literature have been mixed. One of the earliest studies found no association between age and intensity of expression as predicted (Herba, Landau, Russell, Ecker & Phillips, 2006). Alternatively, inclusion of an adult group in a study investigating sensitivity to fear and anger expressions revealed that only adults had significantly greater sensitivity to anger than both children and adolescents, but for fear only differed significantly from children (Thomas *et al.*, 2007). Quantitative differences determining the stage of development when full maturity is reached for each emotional expression can therefore only be established with the inclusion of an adult group, which has not previously been adopted by all studies. A variety of emotional expressions have also been included across behavioral approaches, making some cross-study comparisons difficult. Only one study, Gao and Maurer (2010), has included all six basic emotions in a non-computerized task to investigate sensitivity to expression intensity; however, during the task the expressions were divided into two subgroupings so all six emotions were not presented at once together. Similarly, across studies there has been variation in how developmental age groups are defined, with most studies comparing age groups of between 3 to 5 or more years difference, again making some cross-comparisons of studies difficult. In sum, studies controlling for the continued development of emotional expression recognition from childhood through adolescence into early adulthood remain very limited (Herba & Phillips, 2004). To the best of our knowledge there is only one such developmental study investigating relatively broad age groupings of between 3 to 5 years difference, from age 7 years up to adulthood (Thomas *et al.*, 2007). However, this study was limited to fear and anger expressions only. Further, much of the research focuses on younger age groups, therefore providing only a snapshot of differences at a particular stage of development. Essentially, both empirical limitations leave the question of how the development of facial expression recognition unfolds from early childhood up to adulthood unresolved.

Targeting these prevailing gaps in the literature, the primary aim of this study was to map for the first time the continuous development of facial expression recognition in children aged 5 up to adulthood for each of the six basic emotions and a neutral expression using a psychophysical approach. To the best of our knowledge, this is the first time a psychophysical approach has been used to investigate the development of facial expression recognition. We used the QUEST adaptive staircase procedure (Watson & Pelli, 1983) to establish each participant's recognition threshold for expression discrimination in a signal detection paradigm. The QUEST procedure parametrically manipulated the quantity of face signals available in the stimulus to determine the threshold signal strength at which an expression could be categorized, with lower thresholds indicating more effective discrimination. Based on previous developmental literature, we predicted a general improvement in recognition thresholds with age, and distinct developmental trajectories for each of the expressions, with happiness being recognized most easily at the lowest threshold and fear with most difficulty at the highest threshold.

Methods

Participants

One hundred and sixty individuals participated in the study: 20 adults (M = 21.1 years, 18 females), 60 adolescents: 20 17–18-year-olds (M = 17.7 years, 13 females), 20 15–16-year-olds (M = 15.7 years, 11 females), 20 13–14-year-olds (M = 13.5 years, 11 females), and 80 children: 20 11–12-year-olds (M = 11.5 years, 9 females), 20 9–10-year-olds (M = 9.5 years, 10 females), 20 7–8-year-olds (M = 7.5 years, 10 females) and 20 5–6-year-olds (M = 5.6 years, 10 females). Children were recruited from local schools in the Fribourg and Glasgow regions, and parental consent was obtained for all children under the age of 16. The study was approved by the Department of Psychology Ethics Committee at the University of Fribourg.

Materials

The stimuli consisted of 252 grey-scale images from the KDEF (Lundqvist, Flykt & Öhman, 1998) comprising 36 distinct identities (18 male) each displaying six facial expressions (fear, anger, disgust, happy, sad, surprise) and a neutral expression. Images were cropped around the face to remove distinctive hairstyles using Adobe Photoshop, and were aligned along the eyes and mouth using Psychomorph software (Tiddeman, Burt & Perrett, 2001). The images (256 × 256 pixels) were similarly normalized for contrast and luminance using the SHINE toolbox (Willenbockel, Sadr, Fiset, Horne, Gosselin *et al.*, 2010) in MATLAB 7.10.0 and displayed on an 800 × 600 grey background at a distance of 50 cm subtending $10^{o} \times 14^{o}$ to simulate a natural viewing distance during social interaction (Hall, 1966). The stimuli were presented on an Acer Aspire 5742 laptop using the Psychophysics toolbox (PTB-3) with MATLAB 7.10.0 and QUEST (Watson & Pelli, 1983), a Bayesian adaptive psychometric method (described below) to produce the level of stimulus intensity for each trial. An external USB keyboard was attached to the laptop so the experimenter could key the responses on behalf of the child participants.

Procedure

Before participating, to familiarize the children with the computerized emotion recognition task, each child was shown seven faces expressing the six basic emotions and a neutral expression on individually printed sheets of paper and asked to respond to the question, 'How do you think this person is feeling?' To facilitate the familiarization task for the younger children in particular, the first image presented was always a happy face. If children were unsure of an emotional expression in the familiarization task they were told what the emotion was. The children were then asked if they could repeat this task by looking at images on a computer; however, this time the faces would be slightly hidden or blurred so it might be more difficult to see what the person was feeling, but to please respond as well as they could. Children aged 12 and under responded verbally and the experimenter keyed the response. Children were also told that if they were unsure of an expression, or could not sufficiently see the expression to make a judgement, they could say 'next' and a new face would be presented. Such responses were then coded as 'don't know' by the experimenter. Adolescent and adult participants were told that they would see a series of faces expressing an emotion and were asked to respond as accurately as they could about which emotional expression they saw by pressing the corresponding key on the keyboard. Labels were placed on the bottom row of keys for each of the seven expressions, and on the space bar for 'don't know' responses. Adolescent and adult participants were given as much time as they needed to familiarize themselves with the response keys before beginning the experiment and were told that accuracy not response time was important so to take as much time as needed and to look at the keys if necessary before giving their response.

The experiment began with 14 practice trials to allow participants to become familiar with viewing faces covered with random noise. The transition from practice trials to experiment proper was seamless so the participant was not aware that the initial trials were for practice only. At the beginning of each trial a fixation cross was presented for 500 ms to attract the participant's visual attention, followed by a 500 ms presentation of the face stimulus displayed at the estimated level of signal strength from the QUEST psychometric procedure (described below), directly followed by a mask of random noise (see Figure 1 for an illustrated example of a trial). The emotional expression stimuli were displayed randomly and when the recognition threshold for an expression was obtained (see section below on the QUEST procedure for details), that particular expression was no longer displayed and only images of the remaining expressions were sampled. Keying a response triggered the subsequent trial, so care was required with children to ensure that they were ready for the next stimulus presentation before the response was entered. The number of trials for each participant varied as a function of the QUEST procedure (again described below), so for the youngest children the experiment was paused at roughly midway and continued after a break.

The QUEST Bayesian adaptive psychometric procedure

QUEST is a psychometric function that uses an adaptive staircase procedure to establish an observer's threshold sensitivity to some physical measure of a stimulus, most commonly stimulus strength (Watson & Pelli, 1983). The threshold obtained by the procedure therefore provides a measure of how effectively an observer can discriminate a stimulus. Adaptive staircase procedures obtain the threshold by adapting the sequence of stimulus presentations according to the observer's previous responses. Adaptive staircase methods can therefore be seen as more efficient in determining the observer's perceptual threshold for stimulus detection since the range of stimuli presented is reduced by staying close to the observer's threshold by accounting for their previous responses.

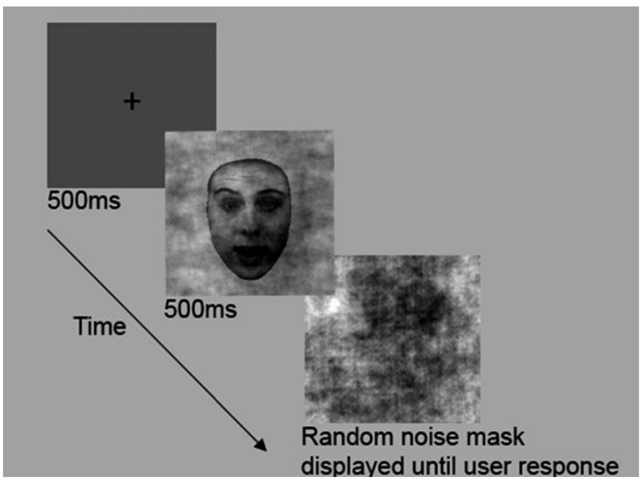

Figure 1 *Example trial. Each trial began with a fixation cross presented for 500 milliseconds, followed by 500 millisecond presentation of a randomly selected face expressing one of six emotional expressions (happiness, surprise, fear, anger, disgust, sadness) or a neutral expression (randomly sampled from the 252 available images) at a signal strength estimated by the QUEST procedure, followed by a random noise mask which remained on the screen until the user provided a response and the next trial was initiated. All images were normalized for contrast and luminance.*

We adopted QUEST for this efficiency as it allowed us to implement a paradigm including all seven expressions at once for the first time in a developmental study. The QUEST threshold-seeking algorithm was implemented in MATLAB 7.10.0 with the Psychophysics Toolbox (PTB-3) to parametrically determine an observer's perceptual threshold for discriminating each of the six emotional expressions and a neutral expression. Adopting a signal-detection approach, QUEST was used to parametrically adapt the signal strength of the grey-scale facial expression images presented to the participant by adding a mask of random noise to the image according to the current signal strength parameter determined by the function, based on the participant's previous performance. If the expression was accurately or inaccurately discriminated on a given trial, then the subsequent signal strength estimate was decreased or increased. The final threshold estimate is determined as the signal strength where the expression is predicted to be discriminated on 75% of trials. In this way equal performance is maintained across observers. Three QUEST procedures were implemented each with different initial stimulus strengths (60%, 40%, and 20%) to prevent possible bias in the final estimate towards the direction of the initial value. The threshold for detecting an expression was therefore the mean of the final estimates from each of the three procedures. The QUEST procedure terminates for an expression after three consecutive correct or incorrect trials in which the signal strength standard deviations are less than 0.025. The threshold is then calculated as the mean stimulus strength of these trials.

Data analyses

Two-way mixed model ANOVA

To investigate the effect of age on emotion recognition thresholds, we performed a mixed model repeated measures ANOVA with emotional expression (7) as the within-subjects factor, and age group (8) as the between-subjects factor.

The threshold estimated by QUEST is the best indicator of performance using an adaptive staircase procedure. The number of trials might not be indicative of performance with this type of procedure, as a short number of trials can be indicative of either a well-recognized expression or a poorly recognized expression that terminated quickly due to consistent inaccurate categorization. Conversely, a long sequence of trials can be indicative of a mixed performance, which has alternated between correct and incorrect responses. For this reason we will not examine the number of trials here.

Generalized Linear Model Regression analyses with bootstrap procedure

In order to characterize the decrease or increase in the amount of information required to accurately categorize an expression across development, we fitted General Linear Models (GLMs) across age groups independently for each emotional expression. For each expression, we sampled with replacement the participants' mean recognition thresholds independently per group. We then computed the trimmed mean (30%) across 20 randomly chosen participants (with replacement), and repeated this procedure 1000 times, leading to 1000 (samples) $\times$ 7 (emotions) $\times$ 8 (age groups) threshold scores. We then used GLM to fit a line across the 8 age groups independently per emotion and sample, thus obtaining 7 $\times$ 1000 fitted linear models. We took the first derivative of each fitted line (which is equivalent to the beta obtained by fitting a GLM with an intercept), resulting in 1000 derivative values. Each derivative thus indicates the rate of decrease/increase in the amount of information required to categorize an emotional expression across age groups. To test whether the increase/decrease across age groups in the amount of information required to categorize an expression was significantly different between emotional expressions we computed 95% confidence intervals (btCI) on the differences (across all pairs of emotions) for our 1000 bootstrapped derivatives. For a given comparison, btCIs non-overlapping with zero indicate that the rate of decrease across development is significantly different.

Similarity matrix and multidimensional scaling analyses

To further characterize the relationship between age and expression recognition we computed a similarity matrix by correlating the average recognition threshold for all expressions across groups. We computed the mean across participants independently for each age group and emotional expression, leading to eight vectors (one per group) of seven expression thresholds. We then iteratively Pearson correlated these vectors across all groups to obtain our similarity matrix. Each value within this matrix thus indicates the similarity of response profiles between two age groups. To clarify which age groups showed closest similarity in response profiles (i.e. calculated by correlating the vector of the mean recognition thresholds for all expressions across two age groups) during development we conducted multidimensional scaling analysis with a metric stress criterion. This produced an unsupervised arrangement (i.e. without presupposing categorical structure) of the age groups according to their response similarity (Torgerson, 1958; Shepard, 1980; Edelman, 1998). Thus age groups placed close together elicited similar response patterns.

Results

Mean expression recognition thresholds across development

The mean age group recognition thresholds for each expression category are shown in Figure 2, which provides a visual re-representation of the thresholds required for expression categorization across age groups. Overall, the mean recognition thresholds improve with age between the youngest and oldest age groups. As predicted, the happy expression had the lowest perceptual threshold across all age groups. This expression could be discriminated at very low signal levels from the youngest age group. Conversely, all age groups showed the highest perceptual threshold for fear. Almost a full strength signal was required by all age groups to categorize this expression. Between the highest and lowest thresholds there was variation across age groups in the ranking of the thresholds for the remaining expressions. Figure 3 also illustrates the mean age group thresholds, line-plotted for each expression.

ANOVA recognition thresholds by age group

The mixed model repeated measures ANOVA examining the effects of age (8 age groups) and emotional expression revealed significant main effects for both age, $F(7, 152) = 11.4$, $p = .000$, and emotional expression, $F(4.46, 678.09) = 303.34$, $p = .006$. The interaction between age group and emotional expression was also significant, $F(42, 152) = 3.76$, $p = .000$, with Greenhouse-Geisser corrections applied to the within-subjects factor.

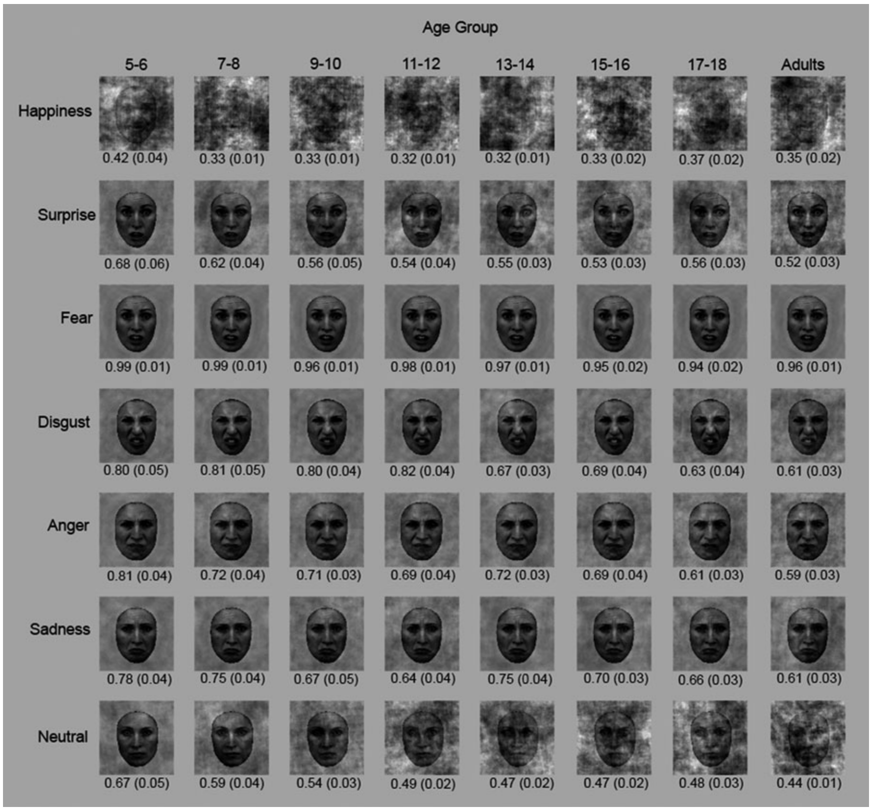

Figure 2 *Mean recognition thresholds across development. Mean recognition thresholds for each expression category per age group. Numbers in parenthesis report the ± standard errors of the mean.*

Generalized Linear Model regression analyses with bootstrap procedure

The General Linear Model regression analyses (Figure 4) revealed a general improvement across development in facial expression recognition for all expressions except fear and happiness. The slope declines from the fitted models illustrate the level of decrease in recognition thresholds with age, and thus the decrease in the amount of information required to categorize an expression. Each emotional expression showed a unique trajectory across development and these trajectories could be more broadly categorized into three groups: expressions that showed a steep improvement in recognition with age up to adulthood – disgust, neutral, and anger; expressions with a more gradual improvement across development – sadness, surprise; and expressions that remained stable from age 5 up to adulthood – happiness and fear. The disgust expression showed the steepest improvement in recognition with age, closely followed by neutral. Alternatively, happiness and fear showed no significant improvement across age with slope derivatives remaining close to zero.

Figure 5a illustrates the boxplots of the means of the 1000 bootstrapped derivatives for each expression. Mean derivatives closest to zero indicate no rate of change in recognition thresholds across age groups. Recognition thresholds for fear and happiness with mean derivatives close to zero therefore did not improve across development. The disgust, neutral and anger expressions with mean derivatives furthest from zero showed the steepest rate of improvement in recognition thresholds across development. The mean derivatives for sadness and surprise fall between the no rate of change fear and happiness expressions and the expressions showing the steepest rate of improvement across development: disgust, neutral and anger.

Figure 5b shows the 95% confidence intervals (btCI) on the differences across all pairs of emotions for the 1000 bootstrapped derivatives. High and low CIs non-overlapping with zero indicate significant differences between the mean slopes of a given pair of derivatives; the 95% btCIs show that the rate of decrease across development for both fear and happiness differed significantly from all other expressions.

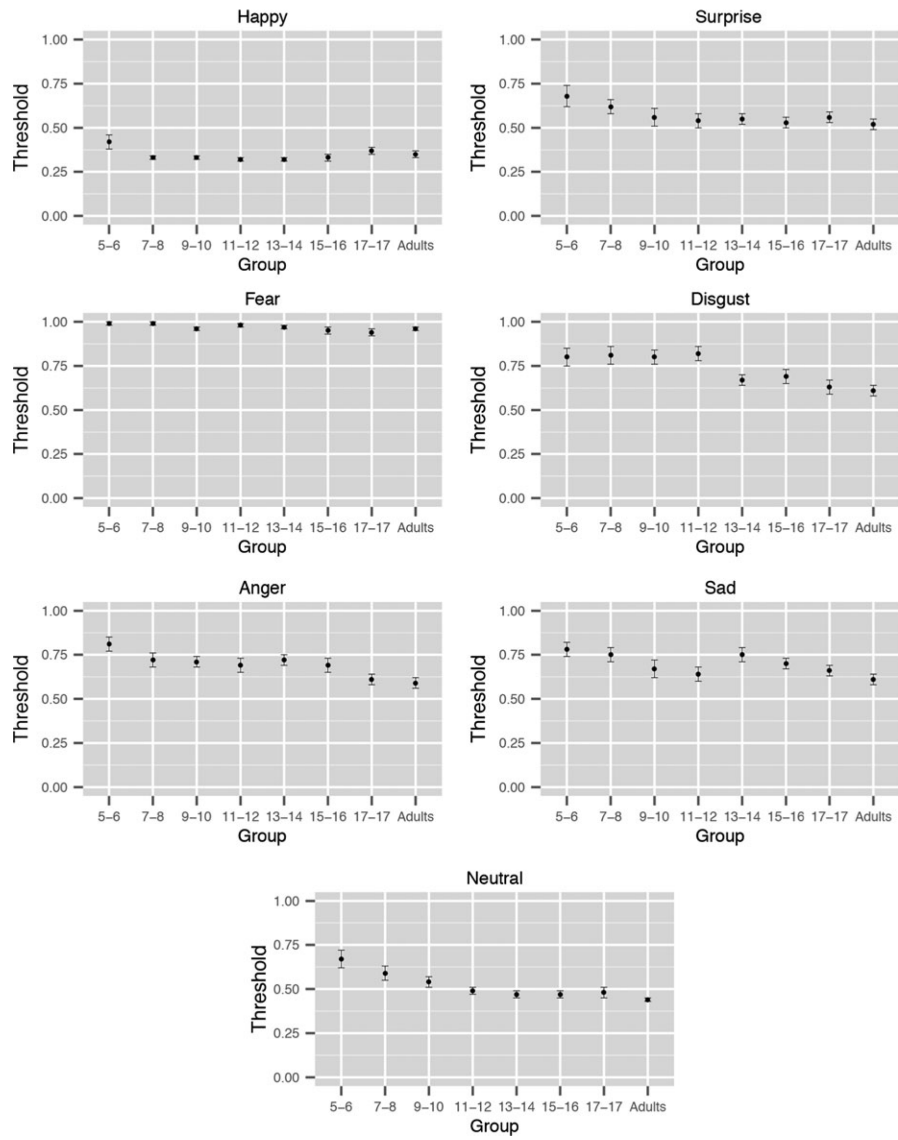

Figure 3 *Age group mean recognition thresholds plotted per facial expression of emotion. Error bars report the ± standard errors of the means.*

Similarity of emotion recognition thresholds across development

The correlation of group means for all seven expressions between age groups is illustrated with a similarity matrix in Figure 6a. Recognition thresholds were most similar between age groups closest in proximity. More broadly, the youngest age groups up to age 12 correlated well, as did the older age groups from age 13 up to adulthood. The multidimensional scaling analysis (Figure 6b) verified which age groups across development showed the most similar response profiles in overall mean recognition scores; the mean squared distances of age groups 5 to 12 clustered together showing similar overall response patterns, as did the age groups from 13 up to adulthood, suggesting that there are two main phases during development in the recognition of facial expressions of emotion.

Response biases for emotion categories

Finally, to examine response biases we calculated confusion matrices for each expression and age group (Figure 7). For all age groups, fear was the most commonly confounded expression, as shown in the top right-hand corner of the confusion matrices for each age group. Fear had the highest confusion rate with surprise, reaching up to 40% in the 11–12 age group. The

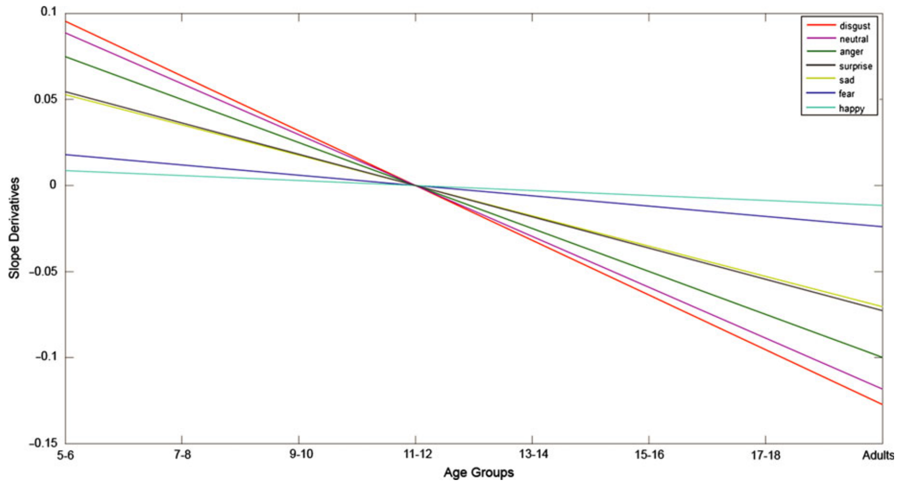

Figure 4 *Linear Models: recognition thresholds across development. Slope decline of the fitted General Linear Model indicates the level of decrease in recognition thresholds with age. The derivatives were centred on the 11–12 age group for visualization purposes. Disgust, neutral, and anger expressions show the greatest level of improvement in recognition with age. Recognition for sadness and surprise improves more gradually with age whereas happiness and fear remain stable.*

confusion rate for fear with disgust also increased between the ages of 15 and 18 but remained lower than that of fear and surprise. The second most commonly confounded expressions were disgust and anger, reaching 30% for 7–8-year-olds and to a lesser extent disgust with sadness, reaching 21% for 9–10-year-olds. Lastly, sadness was most frequently confounded with the neutral expression across all age groups, with rates resting between 15 and 20%.

Discussion

Our results provide a fine-grained mapping of the development of facial expression recognition for all six basic emotions and a neutral expression throughout childhood into adulthood. Using a novel psychophysical approach that maintains equal performance across observers and facial expressions of emotions, we parametrically manipulated the quantity of face signals available to determine an observer's perceptual threshold for each of the six basic emotions and a neutral expression. By controlling for low-level properties such as contrast and luminance, we precisely estimated the quantity of signal necessary to achieve effective recognition for all basic facial expressions of emotion for the first time with young children up to adulthood. The precision and novelty of this approach therefore offer new insight into the understanding of how the development of facial expression recognition unfolds across development.

Overall, recognition accuracy improved with age for all expressions except fear and happiness, for which all age groups including the youngest remained within the adult range. Across development, happiness was the easiest expression to recognize as it was correctly categorized with minimum signals, whereas fear was the most difficult requiring maximum signals. This result confirms the particular status of both facial expressions of emotion, suggesting that the coding for these expressions is already mature at 5 years old. While fear and happiness were the most difficult and easiest expressions to recognize, the developmental profile of each expression was unique. Unique developmental trajectories for the recognition of individual facial expressions of emotion have been reported in the literature previously and our results provide further evidence of this uniqueness (Boyatzis, Chazan & Ting, 1993; Vicari *et al*., 2000; Herba & Phillips, 2004). In particular, our findings characterize the unique trajectories in recognition of the six basic emotions into three distinct groupings: expressions that show a steep improvement in accuracy with age up to adulthood – disgust, neutral, and anger; expressions with a more gradual improvement across development – sadness, surprise; and expressions that remain stable from age 5 up to adulthood – happiness and fear. Two main stages in the development of facial expression recognition were also identified. In the first stage, between the ages of 5 and 12 years, recognition thresholds across expressions followed a similar response profile and developed progressively. The second stage of development began with the onset of adolescence and

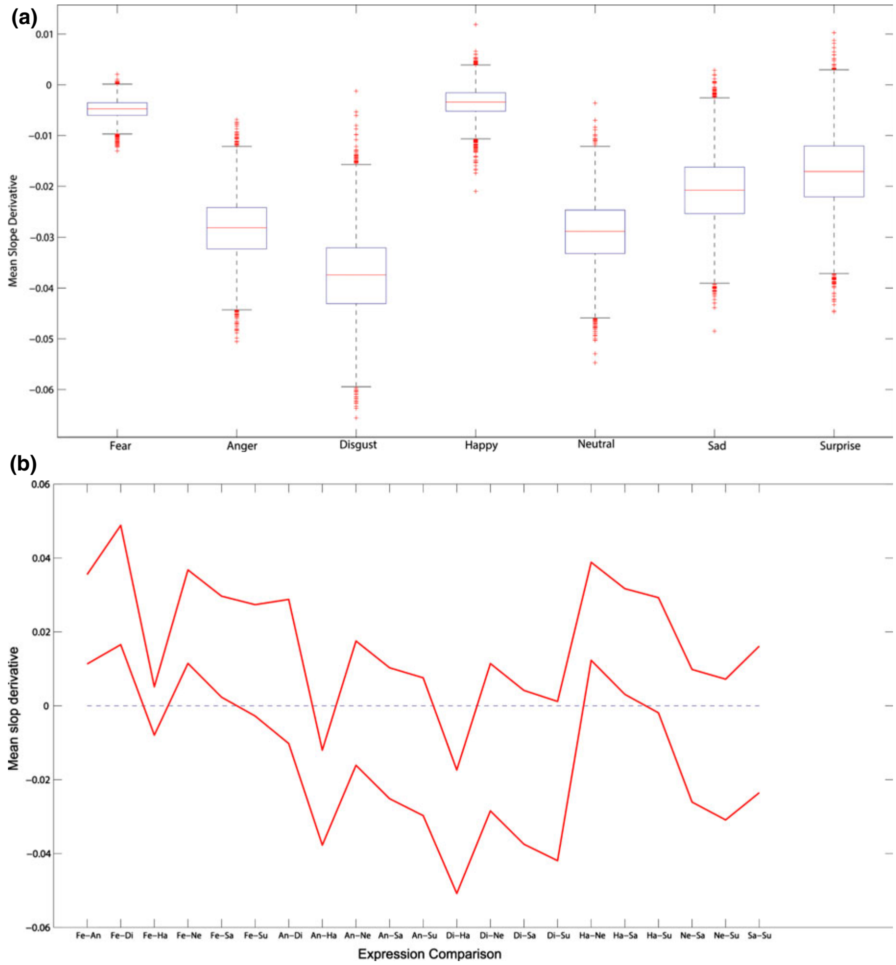

Figure 5 *(a) Boxplot of the derivatives from the bootstrap populations. Boxplots of the derivatives from the bootstrap populations of the lines fitted across age groups independently per expression recognition threshold. The central mark reports the median of the distribution, the edges of the box are the 25th and 75th percentiles (i.e. interquartile range – iqr), the whiskers extend to a maximum of 1.5 times the length of the iqrs. Values falling 1.5 times outside the iqr are considered outliers and plotted as red crosses ('+'). Smaller values indicate greater slope decline and therefore greater improvement in recognition with age. (b) 95% bootstrapped confidence intervals of the mean slope derivatives for all possible expression comparisons. The upper and lower red lines depict the high and low 95% bootstrap confidence intervals (CIs), respectively. High and low CIs that both fall on either side of zero, represented by the dashed line, indicate significant differences between mean slope derivatives across two conditions. Both fear and happiness differ significantly from all the other facial expressions of emotion.*

continued up to adulthood as recognition thresholds across expressions during this stage also showed similar response profiles. It is worth noting that such response patterns were related to similar mean recognition thresholds for all of the expressions (see Figure 2), ruling out the possibility that the high correlation values were a result of significantly lower overall thresholds between two age groups. Within the second stage described here, our data do not clearly distinguish whether there is an additional period during early adolescence where the overall threshold diverges from the oldest three age groups. Further studies including more measures are necessary to clarify this pattern during early adolescence. Altogether, our data show that the recognition of facial expressions of emotion does not follow a unique monotonic dynamic throughout development.

Emotional expressions with a steep improvement in recognition across development: disgust, neutral, anger

Within the first grouping of expressions showing a steep improvement in performance with age, anger has similarly been found to show a sharp increase during

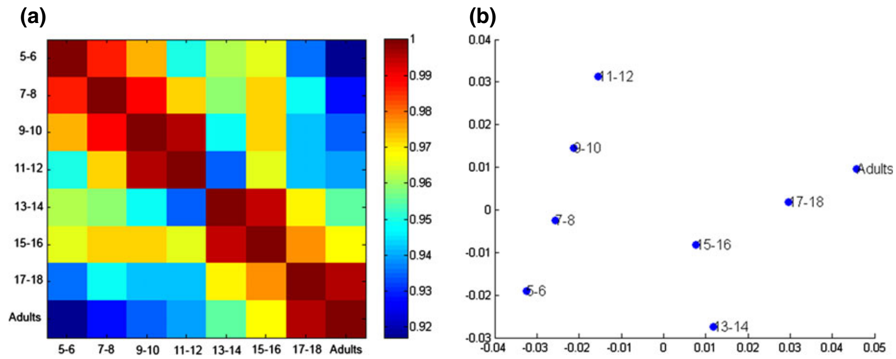

Figure 6 *Similarity matrix of recognition thresholds by age group. (a) Similarity matrix (i.e. matrix of correlated pair values) of the mean group threshold profiles, for all six expressions plus neutral. Dark red indicates high similarity, and the values on the diagonal are of 1. (b) Multidimensional Scaling Analysis: Euclidean distances of overall mean group thresholds.*

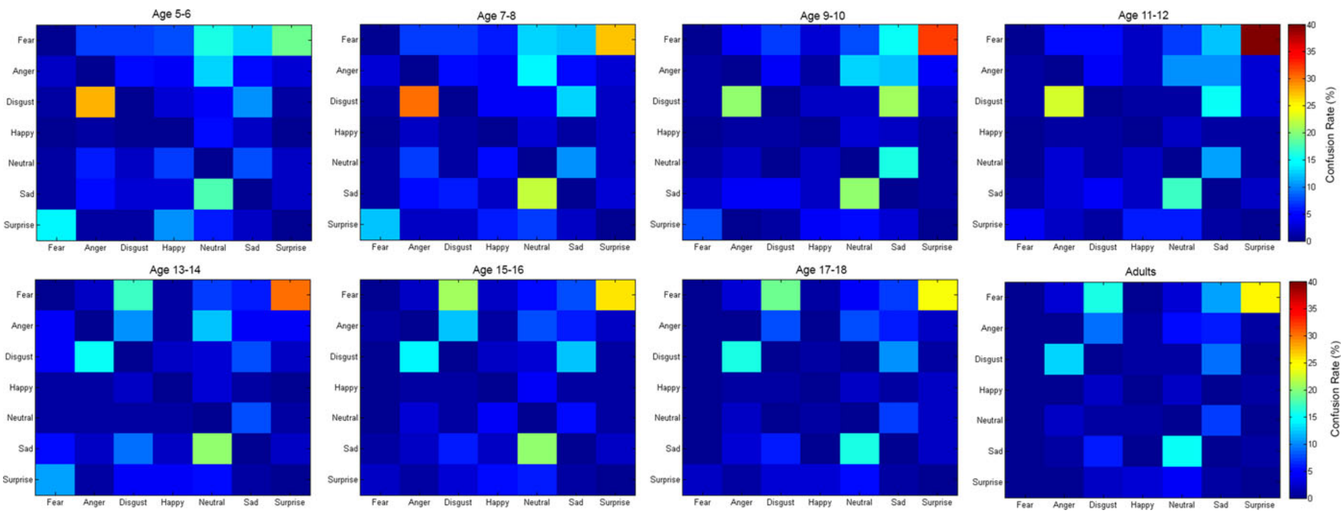

Figure 7 *Facial expression of emotion categorization errors – confusion matrices. Response biases for expression categorization across age groups (%).*

development in several other studies (Montirosso *et al.*, 2010; Gao & Maurer, 2010; Thomas *et al.*, 2007; Vicari *et al.*, 2000), although its comparative trajectory with all seven expressions at once has not been identified previously as we show here. In studies examining recognition performance as a function of expression intensity, somewhat closer to the psychophysical methodology employed here, all have shown a sharper increase in recognition of anger but not disgust with the exception of Herba *et al.* (2006) who report steeper improvements for both disgust and fear but not anger. Thomas *et al.* showed a marked increase in sensitivity to anger from adolescence to adulthood, but examined only fear and anger expressions. While the study did not investigate the neural underpinnings of this result directly, Thomas *et al.* suggest that later development in the recognition of anger fits with neurological evidence as the PFC continues to develop throughout adolescence, with the orbitofrontal cortex in particular being implicated in anger recognition (Murphy, Nimmo-Smith & Lawrence, 2003). In addition to neurobiological accounts for later maturation of anger recognition, the effect of experience has also been evidenced as children growing up in hostile environments show higher accuracy for anger than typically developing children (Pollak & Sinha, 2002). More broadly, cultural differences have been shown in face recognition (Blais, Jack, Scheepers, Fiset & Caldara, 2008) and the expectations of facial expression signals and how these signals are decoded in adult populations, so socio-cultural experience also impacts recognition of expressions (Jack, Caldara & Schyns, 2012a; Jack, Blais, Scheepers, Schyns & Caldara, 2009; Jack, Garrod, Yu, Caldara & Schyns, 2012b). Here, younger children showed comparatively more difficulty in recognizing anger than the expressions showing more gradual trajectories, and were more likely to confound anger with a neutral expression as opposed to disgust as the four oldest age groups did. Anger recognition did not reach adult-like performance until the oldest adolescent age group which, coupled with low miscategorization rates across age, suggests that the late maturation of anger recognition could reflect accumulated exposure to this expression which is frequently masked in various social contexts (Underwood, Coie & Herbsman, 1992).

Previously reported trajectories for disgust recognition have been mixed. Most recently, an emotional intensity study found that accuracy for disgust remained at a similar level in children between the ages of 5 and 10 years; however, disgust was tested alongside surprise and fear expressions only, so it is possible that its distinctness from these other expressions could have contributed to the stable performance (Gao &Maurer, 2010). An emotion intensity study using dynamic stimuli similarly did not find an age-related improvement for disgust between preschool and adolescence, but found that anger improved consistently from school age to late adolescence as we report (Montirosso *et al*., 2010). Conversely, an earlier study reports a steep developmental improvement on a labeling task for disgust in children aged 5 to 10, but a ceiling effect for disgust on a matching task using the same stimuli (Vicari *et al*., 2000). Since the visuo-spatial configuration of disgust is very distinctive, the authors suggest that the steeper developmental improvement in the labeling task occurs because of greater lexico-semantic abilities in older children. However, when tested directly a more recent study found that verbal ability does not significantly impact FER whereas labeling ability does (Herba, Benson, Landau, Russell, Goodwin *et al*., 2008). We do not attribute the steep improvement in disgust found here to greater labeling ability in older children since we accepted responses from younger children such as 'he doesn't like it' or more simply a 'yuck' noise as accurate labels of disgust. Considering the more stable performance in disgust recognition found in the emotion intensity studies described above, methodological differences possibly account for this. Notably, the threshold obtained by this paradigm was a measure that was adapted from the observer's previous response accuracy. Previous studies have established intensity measurement increments a priori rather than on a trial-by-trial basis, and these increments have tended to be large so can lack the sufficient sensitivity to identify developmental differences in emotion processing where an adaptive measure permits greater sensitivity. Moreover, it is important to distinguish that while both paradigms use techniques to parametrically reduce the strength of the original expression to establish an observer's recognition sensitivity, one provides a measure of the intensity at which an expression can be recognized, while the other identifies the quantity of information required for accurate expression recognition.

Lastly, neutral, the final expression of the steep increase with age category, was not included in any of the emotional intensity studies previously discussed here as a distinct emotion category since intensity increments are defined by morphing neutral and emotional expressions together. In general, neutral expressions have been under-investigated in behavioral studies so very little is known about how neutral expressions are perceived during childhood. An early review states that children have difficulty recognizing neutral expressions, and to our knowledge no recent behavioral studies have addressed the development of this expression specifically (Gross & Ballif, 1991). Our finding of a steep increase in improvement between the youngest and oldest age groups accords with this reported early difficulty and could be explained by a general bias to attend more to emotive faces throughout our social experiences (Leppanen & Nelson, 2009).

Emotional expressions with a gradual improvement in recognition across development: sadness and surprise

Sadness showed a gradual improvement in recognition across development and followed a similar trajectory to surprise. Generally, corresponding with our findings, children have been shown to perform well in recognizing sadness (Herba & Phillips, 2004; Widen, 2013). Several studies have shown that children aged 5 to 6 years do not perform as well as older children or adults, which accords with the slower developmental trajectory reported here (Vicari *et al*., 2000; Gao & Maurer, 2010; Montirosso *et al*., 2010), and this trend has been more frequently shown in studies of emotion intensity, with the exception of Gao and Maurer (2009) who found that children as young as 5 can recognize expressions of sadness as accurately as adults. Previous studies have shown that sadness is frequently confounded with fear, disgust, or neutral expressions, but have not tested all seven expressions together at once so confusion rates varied according to which expressions sadness was categorized alongside. We show that sadness was most frequently confounded with neutral expressions across all age groups. To establish whether this miscategorization is simply due to closer similarity in the facial configurations of these two expressions, further studies are needed to determine information use that leads to both accurate and inaccurate categorization across development, as a reduction in miscategorization with age is also shown here.

Surprise similarly showed a more gradual improvement in recognition accuracy across development. Of the emotion intensity studies that we have focused on because of their closer similarity to the psychophysical method adopted here, similarly to the neutral expression, development in the recognition of surprise is not well documented. Inconsistency across the range of expressions tested at this stage of development was one of the motivations for conducting this study with all six basic emotions and a neutral expression. More generally, other methodologies have shown that surprise is recognized at a later stage of development than other expressions (Herba & Phillips, 2004; Widen, 2013); however, we found that when all expressions are compared together, the developmental trajectory for surprise is more gradual, with younger children performing well in recognition of surprise but at the same time showing higher confusion of surprise with fear than older age groups.

Emotional expressions that remained stable from early childhood: happiness and fear

Robust recognition of happiness from an early age was demonstrated by this paradigm as even with a very rapid presentation time of 500 milliseconds in comparison to previous developmental studies, and distortion of the emotional expression with a random noise mask, performance for happiness was highest and remained similar across age groups. Other studies have similarly found that children as young as 5 can recognize a happy expression as well as adults, and that happiness is the first expression to be accurately recognized (Gao & Maurer, 2009; Herba & Phillips, 2004; Gross & Ballif, 1991). As mentioned above, of the emotional intensity studies cited here, Herba *et al*. (2006) found a steeper developmental trajectory in the recognition of fear than we report, but the majority of studies showed findings consistent with a more gradual improvement (Thomas *et al*., 2007; Gao & Maurer, 2009, 2010). However, previous emotional intensity studies have shown higher performance in recognition of fear at lower levels of intensity than we report, where almost maximum signals were required across development to categorize fear. While intensity and signal strength provide distinct measures, the high signal strength required here can be explained by the high variance in miscategorization rates for fear across development which prevented lower threshold rates from being achieved. Such variance could not be achieved in previous studies as they have not included all seven expressions simultaneously. Fear was most consistently miscategorized as surprise at the highest rate across development of between 22 to 37%, and variability in miscategorizations was much greater compared to other expressions as confusions were found with all other expressions except happiness across age groups. The high confusion rate for fear and high variability of these confusions indicates that below a full signal level information was insufficient to categorize fear.

Fear as the most difficult or one of the most difficult expressions to accurately recognize with static images is consistently reported in both the developmental behavioral literature (Gross & Ballif, 1991; Herba & Phillips, 2004; Widen, 2013) and the literature on adults (Rapcsak, Galper, Comer, Reminger, Nielsen *et al*., 2000; Calder, Keane, Manly, Sprengelmeyer, Scott *et al*., 2003); however, this difficulty is frequently juxtaposed with the evolutionary argument that accurate recognition of fear is critical to our survival in comprehending environmental threats. Fear is perhaps the strongest multisensory expression; for instance, people may shout when expressing fear. Consequently, our results and previous results showing difficulty in the categorization of fear suggest that this expression requires additional information to be effectively recognized, when presented as a static image in conjunction with several expressions. Additional cues from other modalities, and body posture or context, may enable more consistent recognition (Aviezer, Hassin, Ryan, Grady, Susskind *et al*., 2008). Overall, our data show that fear and happiness share a special status in the framework of facial expression recognition as the coding for these expressions is already mature by 5 years of age.

Lastly, whether the perceptual mechanisms governing facial expression recognition are holistic or feature based, or whether each type of processing plays a differential role according to particular facial expressions is still debated in the adult literature (Beaudry, Roy-Charland, Perron, Cormier & Tapp, 2014). Although this question is outside the scope of our study, it should be acknowledged that if holistic processing was affected by the use of noise to control for the quantity of signal, then all expressions were equally affected by this manipulation. Therefore, this potential impairment to holistic processing does not straightforwardly account for the differences in recognition thresholds across expressions and age groups. Our results and previous work (Jack *et al*., 2009) would instead favor a feature-based processing account for facial expression recognition. However future developmental studies are necessary to directly address this issue, with paradigms controlling for facial feature information and metrics.

Conclusion

Our data provide a fine-grained mapping of the development of facial expression recognition for the six basic emotions and a neutral expression in children aged 5 up to adulthood. The novel psychophysical approach offers new insight into the understanding of how facial expression recognition unfolds, firstly by characterizing the developmental trajectories of expression recognition into three distinct groupings: expressions that show a steep improvement in accuracy with age up to adulthood – disgust, neutral, and anger; expressions with a more gradual improvement across development – sadness, surprise; and expressions that remain stable from age 5 up to adulthood – happiness and fear; and secondly by identifying two main stages in the development of facial expression recognition: from age 5 to 12, and 13 up to adulthood. These insights have implications for caregivers and educators working daily with young children, particularly for expressions showing a steep improvement with age such as anger, as we show here that in early childhood anger is not easily recognized. Lastly the fine-grained scale of this approach in mapping the development of FER provides a benchmark for thresholds in typically developing children and offers a novel tool to measure impairments to individual facial expressions in developmental clinical populations, such as children with autism spectrum disorders or social behavioral disorders.

Acknowledgements

We would like to thank all of the children who participated in our study from the following schools in Switzerland and Scotland: Ecole Primaire de Marly Cité, Swinton Primary School, Larbert High School, Ecole Enfantine et Primaire de Marly Grand Pré, Cycle d'Orientation de Pérolles and Ecole de Culture Générale Fribourg. We would also like to thank the teachers for their patience with classroom interruptions, and Head Teachers, Claude Meuwly, Michelle Wright, Dominique Brulhart; Principal Teachers, Paul Rodger, Claude Fragnière; and Technician Christophe Pochon, for organizing the logistics of our visits, which ensured that everything ran smoothly. We would also like to thank Stephanie Cauvin, Bérangère Prongue, Elena Bayboerek, Sonia Pupillo, and Sarah Degosciu for their help with data collection. This study was supported by the National Center of Competence in Research (NCCR) Affective Sciences financed by the Swiss National Science Foundation (no. 51NF40-104897).

References

Aviezer, H., Hassin, R.R., Ryan, J., Grady, C., Susskind, J. *et al.* (2008). Angry, disgusted, or afraid? Studies on the malleability of emotion perception. *Psychological Science*, **19**, 724–732.

Beaudry, O., Roy-Charland, A., Perron, M., Cormier, I., & Tapp, R. (2014). Featural processing in recognition of emotional facial expressions. *Cognition and Emotion*, **28**, 416–432.

Blais, C., Jack, R.E., Scheepers, C., Fiset, D., & Caldara, R. (2008). Culture shapes how we look at faces. *PLoS ONE*, **3**, e3022.

Boyatzis, C.J., Chazan, E., & Ting, C.Z. (1993). Preschool children's decoding of facial emotions. *Journal of Genetic Psychology*, **154**, 375–382.

Calder, A.J., Keane, J., Manly, T., Sprengelmeyer, R., Scott, S. *et al.* (2003). Facial expression recognition across the adult life span. *Neuropsychologia*, **41** (2), 195–202.

Edelman, S. (1998). Representation is representation of similarities. *Behavioral and Brain Sciences*, **21**, 449–498.

Gao, X., & Maurer, D. (2009). Influence of intensity on children's sensitivity to happy, sad, and fearful facial expressions. *Journal of Experimental Child Psychology*, **102**, 503–521.

Gao, X., & Maurer, D. (2010). A happy story: developmental changes in children's sensitivity to facial expressions of varying intensities. *Journal of Experimental Child Psychology*, **107**, 67–86.

Gross, A.L., & Ballif, B. (1991). Children's understanding of emotion from facial expressions and situations: a review. *Develomental Review*, **11**, 368–398.

Hall, E. (1966). *The hidden dimension*. Garden City, NY: Doubleday.

Herba, C.M., Benson, P., Landau, S., Russell, T., Goodwin, C. *et al.* (2008). Impact of familiarity upon children's developing facial expression recognition. *Journal of Child Psychology and Psychiatry*, **49**, 201–210.

Herba, C.M., Landau, S., Russell, T., Ecker, C., & Phillips, M.L. (2006). The development of emotion-processing in children: effects of age, emotion, and intensity. *Journal of Child Psychology and Psychiatry*, **47** (11), 1098–1106.

Herba, C., & Phillips, M. (2004). Annotation: Development of facial expression recognition from childhood to adolescence: behavioural and neurological perspectives. *Journal of Child Psychology and Psychiatry*, **45**, 1–14.

Izard, C., Fine, S., Schultz, D., Mostow, A., Ackerman, B. *et al.* (2001). Emotion knowledge as a predictor of social behavior and academic competence in children at risk. *Psychological Science*, **12**, 18–23.

Jack, R.E., Blais, C., Scheepers, C., Schyns, P.G., & Caldara, R. (2009). Cultural confusions show that facial expressions are not universal. *Current Biology*, **19** (18), 1543–1548.

Jack, R.E., Caldara, R., & Schyns, P.G. (2012a). Internal representations reveal cultural diversity in expectations of facial expressions of emotion. *Journal of Experimental Psychology: General*, **141** (1), 19–25.

Jack, R., Garrod, O., Yu, H., Caldara, R., & Schyns, P.G. (2012b). Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences, USA*, **109** (19), 7241–7244.

Johnston, P.J., Kaufman, J., Bajic, J., Sercombe, A., Michie, P.T. *et al.* (2011). Facial emotion and identity processing development in 5 to 15 year old children. *Frontiers in Psychology*, **2** (26), 1–9.

Leppanen, J.M., & Nelson, C.A. (2009). Tuning the developing brain to social signals of emotions. *Nature Reviews Neuroscience*, **10**, 37–47.

Lundqvist, D., Flykt, A., & Öhman, A. (1998). *The Karolinska Directed Emotional Faces.* Karolinska Institute (Ed.). Stockholm.

Mancini, G., Agnoli, S., Baldaro, B., Bitti, P.E., & Surcinelli, P. (2013). Facial expressions of emotions: recognition accuracy and affective reactions during late childhood. *Journal of Psychology*, **147**, 559–617.

Mondloch, C.J., Geldart, S., Maurer, D., & Le Grand, R. (2003). Developmental changes in face processing skills. *Journal of Experimental Child Psychology*, **86**, 67–84.

Montirosso, R., Peverelli, M., Frigerio, E., Crespi, M., & Borgatti, R. (2010). The development of dynamic facial expression recognition at different intensities in 4- to 18-year-olds. *Social Development*, **19**, 71–92.

Murphy, F.C., Nimmo-Smith, I., & Lawrence, A.D. (2003). Functional neuroanatomy of emotions: a meta-analysis. *Cognitive, Affective, & Behavioral Neuroscience*, **3**, 207–233.

Pollak, S.D., & Sinha, P. (2002). Effects of early experience on children's recognition of facial displays of emotion. *Developmental Psychology*, **38**, 784–791.

Rapcsak, S.Z., Galper, S.R., Comer, J.F., Reminger, S.L., Nielsen, L. *et al.* (2000). Fear recognition deficits after focal brain damage. *Neurology*, **54**, 575–581.

Shepard, R.N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science*, **210**, 390–398.

Thomas, L.A., De Bellis, M.D., Graham, R., & LaBar, K.S. (2007). Development of emotional facial recognition in late childhood and adolescence. *Developmental Science*, **10** (5), 547–558.

Tiddeman, B., Burt, M., & Perrett, D. (2001). Prototyping and transforming facial textures for perception research. *IEEE Computer Graphics and Applications*, **21**, 42–50.

Torgerson, W.S. (1958). *Theory and methods of scaling*. New York: Wiley.

Underwood, M.K., Coie, J.D., & Herbsman, C.R. (1992). Display rules for anger and aggression in school-age children. *Child Development*, **63**, 366–380.

Vicari, S., Snitzer Reilly, J., Pasqualetti, P., Vizzotto, A., & Caltagirone, C. (2000). Recognition of facial expression of emotions in school-age children: the intersection of perceptual and semantic categories. *Acta Pediatrica*, **89**, 836–845.

Watson, A.B., & Pelli, D.G. (1983). QUEST: a Bayesian adaptive psychometric method. *Perception & Psychophysics*, **33** (2), 113–120.

Widen, S.C. (2013). Children's interpretation of facial expressions: the long path from valence-based to specific discrete categories. *Emotion Review*, **5**, 72–77.

Willenbockel, V., Sadr, J., Fiset, D., Horne, G.O., Gosselin, F. *et al.* (2010). Controlling low-level image properties: the SHINE toolbox. *Behavior Research Methods*, **42** (3), 671–684.

Received: 21 March 2014
Accepted: 16 October 2014

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1: Raw Scores