Journal of Experimental Child Psychology 174 (2018) 41–59

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

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

Quantifying facial expression signal and intensity use during development

Helen Rodger *, Junpeng Lao, Roberto Caldara *

Eye and Brain Mapping Laboratory (iBMLab), Department of Psychology, University of Fribourg, 1700 Fribourg, Switzerland

A R T I C L E   I N F O

Article history:
Received 19 October 2017
Revised 10 May 2018

Keywords:
Facial expressions
Emotion recognition
Emotion
Development
Psychophysic
Bayesian modelling

A B S T R A C T

Behavioral studies investigating facial expression recognition during development have applied various methods to establish by which age emotional expressions can be recognized. Most commonly, these methods employ static images of expressions at their highest intensity (apex) or morphed expressions of different intensities, but they have not previously been compared. Our aim was to (a) quantify the intensity and signal use for recognition of six emotional expressions from early childhood to adulthood and (b) compare both measures and assess their functional relationship to better understand the use of different measures across development. Using a psychophysical approach, we isolated the quantity of *signal* necessary to recognize an emotional expression at full intensity and the quantity of expression *intensity* (using neutral expression image morphs of varying intensities) necessary for each observer to recognize the six basic emotions while maintaining performance at 75%. Both measures revealed that fear and happiness were the most difficult and easiest expressions to recognize across age groups, respectively, a pattern already stable during early childhood. The quantity of signal and intensity needed to recognize sad, angry, disgust, and surprise expressions decreased with age. Using a Bayesian update procedure, we then reconstructed the response profiles for both measures. This analysis revealed that intensity and signal processing are similar *only* during adulthood and, therefore, cannot be straightforwardly compared during development. Altogether, our findings offer novel methodological and theoretical insights and tools for the investigation of the developing affective system.

* Corresponding authors.
E-mail addresses: helen.rodger@unifr.ch (H. Rodger), roberto.caldara@unifr.ch (R. Caldara).

https://doi.org/10.1016/j.jecp.2018.05.005

Introduction

Perceiving the emotions of others is fundamental to our daily interactions from birth, and by adulthood most have developed the capacity to read the emotional cues of others effortlessly. How we become proficient in reading the emotional cues of others is a critical developmental question because it is well recognized that impaired emotion processing has negative consequences on social functioning and well-being at all stages of development (Carton, Kessler, & Pape, 1999; Feldman, Philippot, & Custrini, 1991; Izard et al., 2001; Nowicki & Duke, 1992). Much research has focused on how we recognize emotion from facial expressions because they are one of the most prevalent cues that communicate our internal affective states.

Different behavioral paradigms have been adopted to understand how our ability to process facial expressions of emotion develops by measuring changes in recognition performance across the lifespan. The broad aim of developmental studies of facial expression recognition, therefore, is to chart at which age specific emotions can be accurately recognized. In doing so, trajectories representing typical development can be identified and, consequently, early identification of impairments in emotion processing is possible. However, the application of different methods, varied developmental age groups, and subsets of facial expressions tested do not provide a uniform picture of how this ability unfolds during childhood and makes comparisons across studies and age groups difficult (Herba & Phillips, 2004). For example, to date, much of the developmental research on facial expression recognition has targeted infancy and preschoolers (Mancini, Agnoli, Baldaro, Ricci Bitti, & Surcinelli, 2013; Thomas, De Bellis, Graham, & LaBar, 2007), and few studies address the continued development of facial expression recognition throughout childhood and adolescence up to adulthood (Herba & Phillips, 2004; Rodger, Vizioli, Ouyang, & Caldara, 2015). No study has directly compared facial expression recognition tasks that use emotional expressions of full intensity with those that use morphed expressions of varying emotional intensities.

To address these discrepancies in the literature, our study focused on the continued development of facial expression recognition from school-aged children of 5 years of age up to adulthood. In a previous study, we mapped the development of recognition of six facial expressions of emotion and a neutral expression using a novel psychophysical approach (Rodger et al., 2015). Here, we compared two distinct measures of facial expression recognition using a psychophysical approach to study the continued development of emotion recognition throughout childhood and adolescence. This approach gives a precise measure of recognition performance across development as the quantity of signal (random image noise blended with emotional facial expression images) or intensity (neutral to facial expression image morphs) is parametrically manipulated. The signal condition is comparable to conventional facial expression recognition categorization tasks that use expressions with 100% phase signals, whereas the intensity condition is similar to tasks using parametric morph designs with expressions of different intensities (however, the intensity increments have been predetermined in studies up until now). The methodological novelty of our psychophysical approach consists in increasing the sensitivity for both tasks by determining an unbiased fine-grained threshold for the effective categorization of facial expressions with a response-driven approach. The theoretical novelty lies in our investigation of whether such commonly used paradigms in the literature relate to the same categorization processes across development or not given that up until now they have been considered interchangeable. To the best of our knowledge, a straightforward relationship between these two measures across development has always been assumed but has never been tested empirically.

Common behavioral methods in the study of the development of facial expression recognition

The most common behavioral methods to investigate facial expression recognition during childhood include matching and labeling tasks and studies of expression intensity. Each method has its strengths and aims to uncover specific features of emotion processing at a given stage of development. Comparison of these common methods can reveal what is consistently found for facial expression recognition during a developmental stage and where methodological gaps or inconsistencies exist. After reviewing these common methods, we describe the novel psychophysical approach we applied to investigate the development of facial expression recognition.

Matching and labeling tasks

Facial expression matching tasks have been employed most frequently in developmental studies of the previous decade. Matching tasks require the child to match one image of an expression to another image of an expression or to one image among several images. Studies using matching tasks with two-, three-, or four-alternative forced choices to the target expression have found that recognition performance progressively improves between 4 and 10 years of age (Bruce et al., 2000; Mondloch, Geldart, Maurer, & Le Grand, 2003; Vicari, Snitzer-Reilly, Pasqualetti, Vizzotto, & Caltagirone, 2000). Such tasks also show converging high-level performance by 10 years. Using a two-alternative forced-choice matching task, accuracy had reached nearly 100% by 10 years in a study that consequently classified this type of task as "easy" among the face-processing tasks investigated (Bruce et al., 2000). Similarly, with a three-alternative forced-choice matching task, by 10 years performance was equivalent to that of adults (Mondloch et al., 2003). Therefore, whereas a high level of performance in matching tasks is possible by 10 years, slight modification of this type of task to a simple pointing exercise between pairs of expressions for the target expression led to ceiling-level performance by 6 years (Bruce et al., 2000). An increase in the number of expression choices, with a four-alternative forced-choice matching task, similarly showed that by 10 years performance was high across the expression categories tested (Vicari et al., 2000). Again, modification of this task showed different performance outcomes. When the target expression was covered after 5 s, thereby placing greater demands on memory, performance consequently dipped in this age group (Vicari et al., 2000). Therefore, even for relatively simple tasks, varying task demands alter recognition performance during development, as is acknowledged in the literature (Johnston et al., 2011; Montirosso, Peverelli, Frigerio, Crespi, & Borgatti, 2010; Vicari et al., 2000). The more recent challenge, therefore, has been to find appropriate tasks with sufficient sensitivity for use across development.

Conventionally, matching paradigms have been used to attempt to minimize verbal ability and memory confounds. However, as illustrated above, across the variety of matching paradigms that have employed tasks of increasing or decreasing complexity, high levels of performance have been shown by middle to late childhood, indicating that this type of task does not challenge the maximum capabilities of children at this stage of development. Addressing studies that show little change in accuracy between 7 and 10 years of age, De Sonneville et al. (2002) proposed that speed of responding can provide a more sensitive measure to reveal age-related changes in facial expression processing. They found that the speed of responding greatly improved during this age range, whereas accuracy improvements were small. However, the task used to obtain a speed of response measure was a simple yes/no response to whether a face shows the target expression. Although this paradigm can reveal developmental changes where accuracy measures cannot, there are also several constraints. Because only yes/no response options are possible, accuracy must be significantly greater than the 50% chance level and the number of emotions that can be presented is limited because a speed of response versus length of task trade-off is expected. Information about miscategorizations across emotions similarly cannot be determined.

Alternatively, labeling tasks, another commonly employed method, allow an unrestricted number of response options and analysis of miscategorizations. In a labeling task, the child must select the correct emotion label for the expression presented from several label options. These tasks have traditionally used either forced-choice response categories or free labeling that allows for an unrestricted number of response options and the analysis of miscategorizations. Here, we focused on school-aged children and the use of labeling in intensity studies similar to the method applied here, but we note the body of work using labeling tasks with younger children by Widen and Russell because it comprises one of the few developmental perspectives of how emotion understanding evolves.

In a series of studies using the labeling of both images of facial expressions and emotion stories, Widen and Russell (2003) developed the differentiation model of emotion understanding. The model describes how children initially understand emotions in terms of the broad dimensions of valence (pleasure–displeasure) and arousal (high–low) rather than as the discrete categories that adults come to understand. Gradual understanding of these initial broad dimensions is slowly narrowed conceptually to discrete categories such as anger and disgust. Because this model is developed from the use of different types of labeling tasks including conceptual stories, it is possible that by restricting the stimuli to perceptual images uniquely, the order in which these labels are acquired may change. Moreover, the type of miscategorizations in a perceptual study may also inform how emotion labels are acquired. Basic visual stimuli are used in the current study and similarly in other studies of emotion intensity that have mainly employed labeling tasks. We now focus on intensity studies because intensity comprises one of the measures here. The verbal ability required for labeling tasks limits their use to school-aged and some preschool-aged groups of children. In comparing matching and labeling tasks, it is evident that the balance between sensitivity in the measure and complexity of the task is challenging to assimilate for the accurate assessment of recognition performance across different developmental age groups.

Expression intensity studies

To obtain a more nuanced understanding of the development of facial expression recognition, one approach in more recent behavioral studies has been to vary the intensity of expressions to establish whether older children can recognize more subtle expressions of emotion in comparison with younger children. Such results are anticipated because in daily life we more frequently perceive subtle expressions of emotion. Expression intensity is modified by creating parametric linear blends of emotions called morphs. Typically, morphs are created by blending a percentage of an emotional expression with a percentage of a neutral expression or another emotional expression. Whereas studies employing this method remain few in the developmental literature, the percentage increments to index intensities across studies vary, with 5% or 10% increments most commonly used. Similarly, the morph stimuli used can be static or dynamic.

The results obtained from studies using static morphs have varied, most likely as a consequence of the various levels of intensity increments used and the various age groups and emotions studied. Our study focused on typically developing children, but previous original findings have effectively illustrated the usefulness of the morphing technique by showing the effect of emotional experience on emotional recognition (Pollak & Kistler, 2002), with physically abused children showing greater sensitivity to anger. An early study to use morphs with typically developing children investigated the correspondence between recognition performance and emotion intensity in three age groups (between 4 and 15 years) and four levels of emotional intensity (25%, 50%, 75%, and 100%) but found no association between age and level of intensity, as predicted (Herba, Landau, Russell, Ecker, & Phillips, 2006). Moreover, comparison between intensity levels and emotion categories in the explicit emotion matching task that was used revealed significant differences only between the lowest and highest intensities, suggesting that the increments were too broad to capture differences in the middle range.

In a study of sensitivity to emotion intensity for fear and anger expressions, again across three distinct age groups (children, adolescents, and adults) but with finer increments of intensity (at 11%), participants needed to judge whether the face stimuli expressed a neutral versus angry expression or a neutral versus fearful expression (Thomas et al., 2007). Sensitivity to emotion intensity was measured by comparing the d' average and d' slope across the three age groups and revealed significant differences only between adults and children for fear and between adults and both children and adolescents for anger. Interpreting the results, the authors suggested that there was a marked increase in sensitivity to anger from adolescence to adulthood, whereas sensitivity to fear showed a more gradual incline with age. However, the relatively broad age categories and intensity increments used in this study may have prevented differences from being revealed across the child and adolescent groups.

Finally, two more recent studies investigated sensitivity to emotion intensity using three child age groups from 5 to 10 years of 2- or 3-year intervals with the finest intensity measures to date, 20 levels of 5% increments for each expression studied (Gao & Maurer, 2009, 2010). The studies also included a broader range of expressions and analyzed miscategorizations. Applying the same methodology, both studies investigated children's responsiveness to emotional intensity by calculating a threshold for accurate discrimination of each emotion—happiness, sadness, and fear in the first study, followed by all six basic emotions in the second study. Thresholds were defined as the intensity level at which 50% of the time the expressive face was recognized as a neutral expression and 50% of the time it was recognized as an expressive face. Importantly, this could mean any expression from the emotional expression categories available and not necessarily the correct one because a second measure for misidentification of expression was also recorded. Unlike the previous studies discussed here, the task was not computerized. Instead, children were asked to physically categorize photographs of emotional expressions of varying intensities.

Gao and Maurer (2009) found different developmental patterns for each of the three emotions investigated. The youngest children matched adult sensitivity for both measures of threshold intensity and misidentification of happiness. For sadness, even the oldest children, aged 10 years, were prone to confusing this emotion with fear, and for the fear expression children did not reach adult-like thresholds until 10 years. Gao and Maurer (2010) expanded the number of emotion categories to include all six basic emotions, which were then subdivided into two groupings based on previous findings of the confusability of emotion categories. Therefore, participants completed the recognition task in two blocks, each with distinct emotion categories and not with all six emotions at once. For all age groups, recognition accuracy for happiness reached ceiling-level performance. Between 5 and 10 years, sensitivity to surprise, disgust, and fear improved, and sensitivity to sadness and anger continued to improve into adulthood.

Although this was the first study to include a broader range of emotion categories, several features of this paradigm make it difficult to draw definitive conclusions on sensitivity to emotion intensity in the age groups studied. Primarily, the true threshold applied here was a composite measure; a misidentification measure was calculated separately to the initial threshold measure. Furthermore, emotions could be misidentified only with emotions belonging in the same subgroup; thus, potential misidentifications across all six emotion categories were not possible. A single measure that accounts for miscategorizations could provide greater precision in the understating of facial expression recognition. Finally, here, as with previous studies investigating sensitivity to intensity, the increments were established a priori, so the granularity of the measure can only be as fine as the predefined increments.

A psychophysical approach

Here, we investigated the continued development of facial expression recognition for all six of the basic emotions from early childhood (5 years of age) up to adulthood. We introduced a novel psychophysical method using the QUEST threshold-seeking algorithm (Watson & Pelli, 1983) to obtain a sensitive measure of facial expression recognition performance across the age groups studied. This algorithm identifies an individual's recognition threshold for an expression with a sensitivity of 1% for intensity measures or less for signal measures. The threshold is adapted online during the execution of the experiment. The algorithm, therefore, permits greater sensitivity in the measure of recognition performance, and intensity increments do not need to be defined a priori as with other methods. Our aim was to (a) obtain a precise measure of the quantity of visual information needed to recognize an expression across development and (b) compare two measures of visual information use, *signal* versus *intensity* thresholds, using an experimental design in which each participant is tested under all experimental conditions to better understand the use of different measures in assessing recognition performance across development. We predicted that recognition performance would improve with age for both measures and that this improvement would be distinct for each expression. Based on previous findings, we predicted that happiness would be the easiest expression to recognize across age groups and that fear would be among the most difficult (Herba & Phillips, 2004; Rodger et al., 2015). The QUEST algorithm was used to identify the recognition thresholds for both the signal and intensity measures; however, we had no prediction as to whether one measure would yield higher or lower thresholds, or whether those measures would be significantly related to one another, because no study had previously compared these distinct measures.

Method

Participants

In total, 159 individuals participated in both the signal and intensity conditions. As described below, participants were analyzed on a continuum of age in years. For simplicity, we list the participants by age group. The adult group consisted of 19 participants (M = 24.2 years, SD = 1.8; 10 female). The adolescent group consisted of 60 participants in total: 20 17- and 18-year-olds (M = 17.9 years, SD = 0.65; 17 female), 20 15- and 16-year-olds (M = 16.0 years, SD = 0.73; 12 female), and 20 13- and 14-year-olds (M = 14.0 years, SD = 0.5; 12 female). The child group consisted of 80 participants: 20 11- and 12-year-olds (M = 11.9 years, SD = 0.5; 8 female), 20 9- and 10-year-olds (M = 9.9 years, SD = 0.59; 9 female), 20 7- and 8-year-olds (M = 7.9 years, SD = 0.61; 13 female), and 20 5- and 6-year-olds (M = 5.9 years, SD = 0.56; 13 female). Children were recruited from local schools in the Fribourg area of Switzerland, and parental consent was obtained for all children under 16 years of age. The study was approved by the Department of Psychology ethics committee at the University of Fribourg.

Materials

For the signal condition, the stimuli consisted of 252 grayscale images (256 × 256 pixels) from the Karolinska Directed Emotional Faces (KDEF) database (Lundqvist, Flykt, & Öhman, 1998) comprising 36 distinct identities (18 female), each displaying six facial expressions (fear, anger, disgust, happy, sad, and surprise) and a neutral expression. For the intensity condition, we used eight identities (four female) expressing each of the six basic emotions from the KDEF (Lundqvist et al., 1998) image database. Abrosoft FantaMorph software was used to create morphs of 100 increments for each identity and emotional expression, ranging from a 1% morph of a neutral face and an expressive face up to a 100% expressive face. The total number of images used, therefore, was 4800 (8 identities × 6 expressions × 100 increments). Example stimuli of different expression intensities and signal strengths are shown in Fig. 1. Participants viewed images only at the intensities calculated by the QUEST procedure. All 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). Images were also normalized for contrast and luminance using the SHINE toolbox (Willenbockel et al., 2010) in MATLAB 7.10.0 and displayed on an 800 × 600 Gy background at a distance of 50 cm subtending 10° × 14° 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 with MATLAB 7.10.0 and QUEST (Watson & Pelli, 1983), a Bayesian adaptive psychometric method (described below) to estimate the level of stimuli strength (signal or intensity) for each trial. An external USB keyboard was attached to the laptop so that the experimenter could key the responses on behalf of the child participants.

Procedure

To familiarize the children with the computerized emotion recognition task, each child was shown six faces expressing the six basic emotions on individually printed sheets of paper and were asked to respond to the question, “How do you think this person is feeling?” To facilitate the familiarization

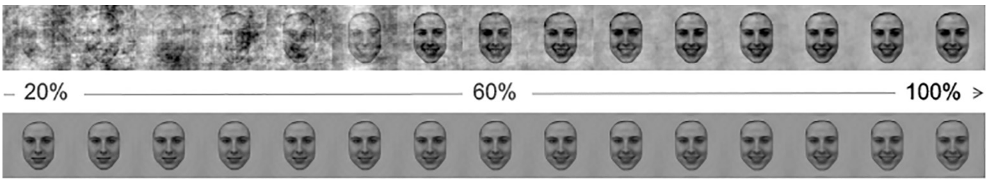

Fig. 1. Example stimuli from the signal and intensity conditions. In this image, the stimuli are shown in increasing 5% increments, starting at 20% signal or intensity.

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. Children were then asked whether they could repeat this task by looking at similar images on a computer screen. For the signal condition, children were told that this time the faces would be slightly hidden or blurred, so that it might be more difficult to see what the person is feeling, but to please respond as well as they could. Because there were six expressions to choose from labeled on six computer keys, children aged 12 years and under responded verbally and the experimenter keyed the responses on their behalf. Children were also told that if they were unsure of an expression or could not sufficiently see the expression to make a judgment, they could say "next" and a new face would be presented. Such responses were coded as "don't know" by the experimenter. Adolescent and adult participants were similarly asked to respond as accurately as they could to how the person in the picture was feeling by pressing the corresponding emotion key labeled on the keyboard. Labels were placed on the bottom row of keys for each of the six expressions and on the space bar for "don't know" or "uncertain" 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 they could take as much time as needed and could look at the keys if necessary before giving their responses.

The experiment began with 6 practice trials to allow participants to become familiar with the computerized task. The transition from practice trials to experiment proper was seamless, so participants were 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 locate the participant's visual attention, followed by a 500-ms presentation of the face stimulus displayed at the signal strength or intensity estimate from the QUEST psychometric procedure (described below), directly followed by a mask of random noise (see Fig. 2 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 "QUEST Bayesian adaptive psychometric procedure" section below 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 short break.

QUEST Bayesian adaptive psychometric procedure

The QUEST procedure as implemented by Rodger et al. (2015) was used. 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 QUEST procedure, therefore, provides a measure of how effectively an observer can discriminate a stimulus. Here, we investigated threshold sensitivity for signal and intensity of expression in two separate conditions across developmental age groups. Adaptive staircase procedures obtain the threshold by adapting the sequence of stimulus presentations according to the observer's previous responses. For example, the stimulus strength becomes weaker or stronger according to the user's history of correct and incorrect responses to a particular stimulus category. Adaptive staircase methods, therefore, can be more efficient in determining the observer's perceptual threshold for stimulus detection because the range of stimuli presented is reduced by staying close to the observer's threshold by accounting for the observer's previous responses.

We adopted QUEST for this efficiency because it allowed us to implement a paradigm including all six expressions at once in a developmental study. The QUEST threshold-seeking algorithm was implemented in MATLAB 7.10.0 with the Psychophysics Toolbox to parametrically determine an observer's perceptual threshold for discriminating each of the six emotional expressions. Adopting a signal detection approach, QUEST was used to parametrically adapt the signal strength of the grayscale facial expression images presented to the participant by adding a mask of random noise to the image corresponding 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

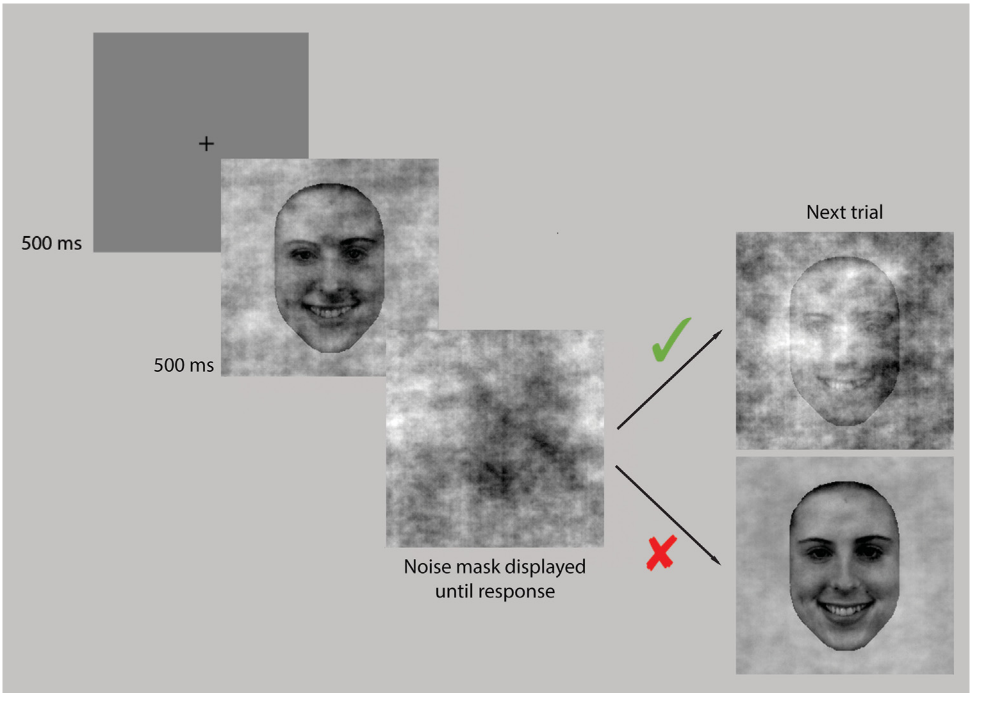

Fig. 2. Example trial from the signal condition. At the beginning of each trial, a fixation cross was presented for 500 ms to locate the participant's visual attention, followed by a 500-ms presentation of the face stimulus displayed at the signal strength or intensity estimate from the QUEST psychometric procedure, directly followed by a mask of random noise until a response was made. Depending on accuracy, the next trial was followed by a face containing more signal (in case of an erroneous response) or less signal (in case of an accurate response).

given trial, then the subsequent signal strength estimate was decreased or increased. Similarly, the intensity of the expression was adapted using neutral expression image morphs from 0% expression (a neutral face) to 100% expression. The final threshold estimate was determined as the intensity or signal strength where the expression was predicted to be discriminated on 75% of trials. In this way, equal performance was maintained across observers. The 75% performance threshold was chosen because it has been conventionally applied in adult face identity and facial expression recognition studies (Gosselin & Schyns, 2001; Schyns, Bonnar, & Gosselin, 2002; Smith, Gosselin, Cottrell, & Schyns, 2005). For the signal condition, three QUEST procedures were implemented, each with a different initial stimulus strength (60%, 40%, or 20%), to prevent possible bias in the final estimate toward the direction of the initial value. For the intensity condition, one QUEST procedure was implemented with an initial expression intensity of 30%. This intensity was selected because by nature 50% intensity denotes an image morph of 50% expression and 50% neutral expression, so the initial value should be below this level of morph. The QUEST procedure terminates for an expression after 3 consecutive correct or incorrect trials in which the intensity or signal strength standard deviations are less than 0.025.

Data analyses

Threshold detection

The participant's recognition threshold for each task is identified as the level of information (intensity or signal) needed to maintain performance at 75%, as quantified by the QUEST procedure. For each expression and participant, the QUEST procedure assumes that the response (in terms of accuracy rate) and the presented signal or intensity follows a psychometric function. Throughout the experiment, this psychometric function is updated and refined for each trial until the end of the experiment. The final threshold estimate is the level of information at which the participant is predicted to maintain 75% performance for expression recognition. This estimate is obtained by computing the mean of the QUEST posterior probability density function (pdf) using the *QuestMean* function from the QUEST toolbox (King-Smith, Grigsby, Vingrys, Benes, & Supowit, 1994; Pelli, 1987), which uses a Weibull psychometric function. In the signal task where multiple QUEST procedures were employed, we computed the average of the threshold estimations as our final estimation for each participant. In our previous study, we used the intensity of the last trial from the QUEST procedure as the threshold estimation (Rodger et al., 2015). However, because some participants cannot achieve 75% accurate identification even when signal or intensity is at the maximum (100%), the previous calculation returns a ceiling value of 1 (which is equal to 100% signal or intensity). This occurs nearly exclusively for fear recognition (see the supplementary figure in Rodger et al., 2015). Instead, here, for the intensity and signal conditions, the threshold estimate now returns values greater than 1 (e.g., 1.112). Because the expression has failed to be categorized at full strength, by using the precise estimate instead, the threshold is no longer constrained by its physical limitation, a value of 1, so a more continuous measure is possible, which gives greater sensitivity to detect potential developmental differences even for expressions that are difficult to recognize across age groups.

Signal and intensity recognition thresholds as a linear function of age per emotional expression

To quantify the relationship between age and emotion recognition performance, we fitted general linear models (GLMs) with age as predictor for each task and expression independently (Rodger et al., 2015). We then compared the regression coefficients between the two tasks for each expression to infer the effect of age (Fig. 4). GLMs were fitted using the *fitlm* function in MATLAB with the default robust option using a bisquare weight function to eliminate the effect of outliers. Hypothesis testing on the model coefficients was corrected for multiple comparisons using a Bonferroni correction.

Response profile analysis

Although the QUEST procedure is efficient in estimating a desired threshold, the returned estimation is a summary statistic that is sufficient only under strong assumptions (e.g., the underlying posterior pdf is parameterized using only the mean). In other words, the uncertainty of the estimation is usually discarded. To fully account for all the information encoded in the response during the QUEST procedure, we applied a Dirichlet–multinomial model to recover the response profile for each participant. This procedure is conceptually described in Fig. 5.

For each participant, we first extracted the raw response vector and the corresponding intensity/signal level for one expression in one task (Step 1 in Fig. 5). We then projected each element in the response vector and its corresponding intensity/signal level into a sparse matrix (Step 2 in Fig. 5). To recover the full response profile from the sparse raw response matrix, we applied a Dirichlet–multinomial model for each intensity/signal level, a probabilistic model widely used to model categorical responses. Here, we assumed that at each intensity/signal level the participant's response to a random stimulus follows a multinomial distribution: $X \sim Multinomial\ (response,\ p)$. The *response* is all six tested expressions plus neutral and the "I don't know" response; p is an eight-element vector (summed to 1) coded for the probability of each response. Moreover, p follows a Dirichlet distribution $p \sim Dirichlet\ (\alpha)$, where α is the concentration parameter of the Dirichlet distribution. To recover the response profile matrix, we started from the lowest intensity/signal level (0%) with a uniform prior for the Dirichlet distribution: a vector of 1 s as α. We applied the Bayes theorem to get the posterior of α and then used the posterior of α as the new prior for the next intensity/signal level (Step 3 in Fig. 5). The Bayesian update procedure is repeated until the highest intensity/signal level (100%) is reached. In this way, the dense matrix representation of the response profile (final output in Fig. 5) is recovered. In the resulting response profile, the value is the concentration parameter α of the Dirichlet distribution in which a higher value relates to a higher concentration in the response probability p in the multinomial distribution.

Specifically, we performed the Bayesian update for each intensity/signal step using PyMC3 with 10,000 Metropolis–Hastings sampling. Instead of sampling from a Dirichlet distribution with the concentration parameter α, we sampled from $\alpha_i \sim Gamma\ (\alpha_I,\ 1)$ and normalized the sum of $[\alpha_1, \ldots, \alpha_k]$ to

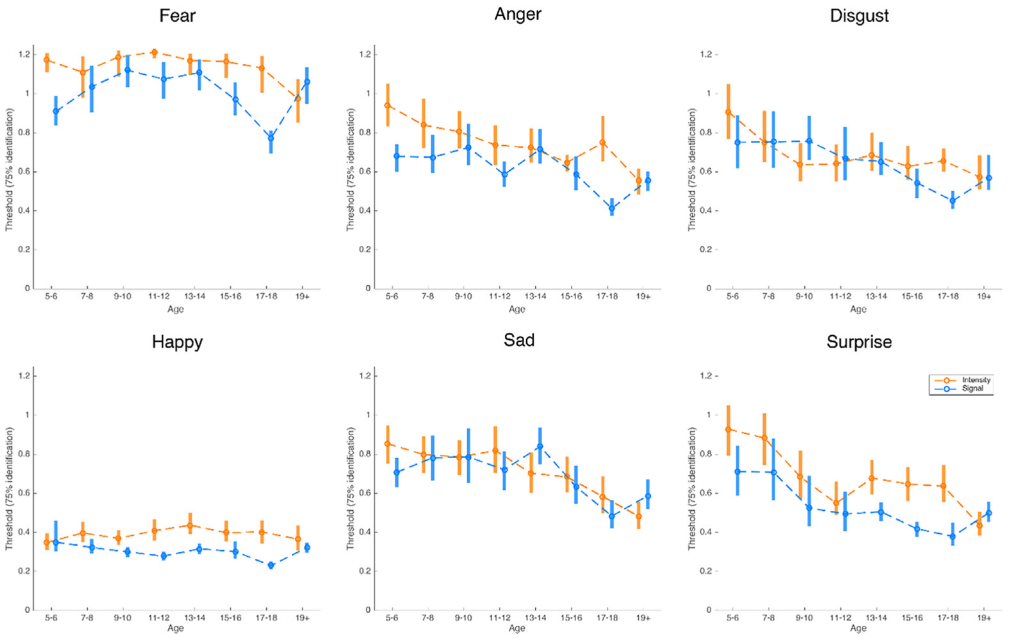

Fig. 3. Age group mean recognition thresholds plotted per facial expression of emotion. Orange lines indicate the intensity task, and blue lines indicate the signal task. Error bars report the 95% confidence intervals. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

1 to get the parameter p for the multinomial distribution. This formulation allowed us to obtain the posterior of the concentration parameter α for the update at the next intensity/signal level.

It is worth noting that the Bayesian update procedure applied here is conceptually similar to fitting a psychometric function independently for each row in the raw response matrix in Fig. 5. However, applying a Dirichlet–multinomial model is more accurate and does not require any collapse of conditions in the computation.

The response profiles for each expression per task across all participants are shown in Fig. 6. To further explore the relationship between the intensity and signal tasks for each participant, we computed the mutual information between the response profiles (using the algorithm in Kinney & Atwal, 2014). Importantly, we included the responses only of the six target expressions (i.e., excluding the rows coded for the neutral expression and "I don't know" response) so that the two tasks were consistent. Robust regression is fitted between the mutual information and age, similarly to as described above, to quantify the effect of age.

Results

Mean expression thresholds across development

Signal

The mean age recognition thresholds and their 95% bootstrapped confidence intervals for each of the expression categories are plotted in Fig. 3. As predicted, happiness was the easiest expression to recognize across age groups, being recognized with the lowest mean thresholds across age groups. In contrast, fear was the most difficult expression to recognize across age groups, having the highest mean thresholds across groups; even at full signal strength, participants generally do not reach the target accuracy (75%) for the fear expression, which resulted in threshold estimations of greater than

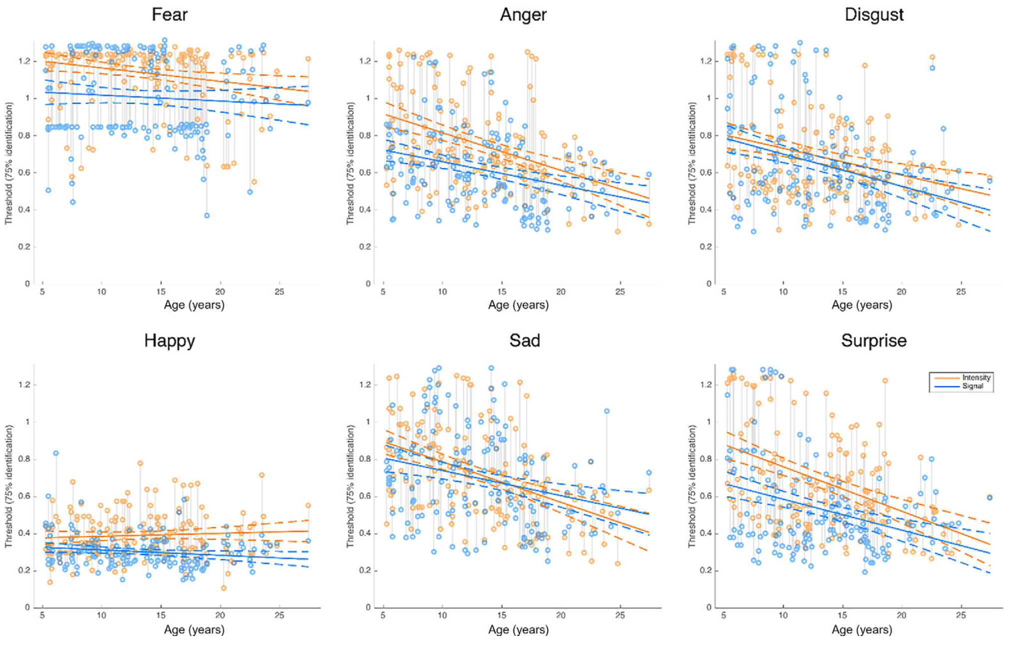

Fig. 4. Individual recognition thresholds plotted as a function of age (*x* axis) per facial expression of emotion. Orange dots indicate the intensity task, and blue dots indicate the signal task. Data from the same participant is linked with a gray line. Longer lines indicate that the thresholds for signal and intensity are not similar and, therefore, are farther apart. Short lines indicate that the thresholds for this individual are similar. Line plots show the results of the linear regression between threshold and age, and dotted lines show the 95% confidence intervals. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

1 by the QUEST algorithm. Across age groups, the rank order of mean expression thresholds between the highest and lowest mean thresholds varied. The mean number of trials for the signal condition was 216.74 (SD = 49.96).

Intensity

The mean recognition thresholds and their 95% bootstrapped confidence intervals for each expression and age group are plotted in Fig. 3. Similar to the signal condition, happiness and fear were the easiest and most difficult expressions to categorize across age groups, respectively, having the lowest mean thresholds for happiness and the highest for fear. Again, as for the signal measure, the majority of participants do not reach the target accuracy (75%) even at full intensity for the categorization of the fear expression. The ranking of mean thresholds between the highest and lowest intensity thresholds again varied across age groups, with no set pattern established across age groups for the remaining expressions. The mean number of trials for the intensity task was 78.06 (SD = 14.15).

Signal and intensity recognition thresholds as a linear function of age per emotional expression

Fig. 4 shows the change in information use across age for each emotional expression. Each individual's threshold for signal and intensity was plotted, with age along the *x* axis. The fitted regression lines for intensity and signal are shown in red and blue, respectively; the dotted line indicates the 95% confidence interval. Overall, a significant decrease in thresholds across age was found for both the signal and intensity measures for four of the six expressions (square brackets show 95% confidence intervals, $p < .05$, Bonferroni corrected): anger (intensity: −.02 [−.0271, −.0135], $t(157) = -5.87$;

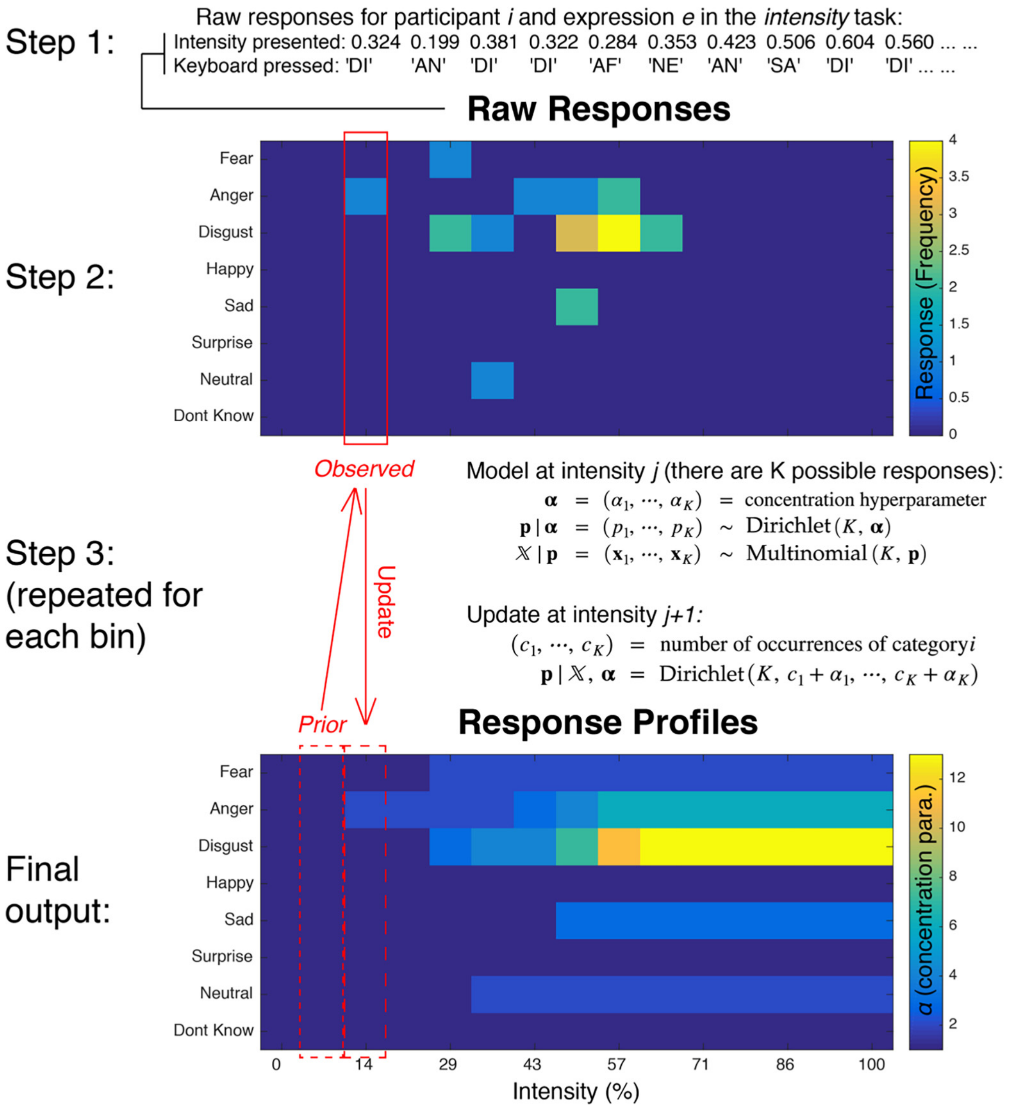

Fig. 5. Response profile analysis for a single participant for one expression (disgust) during the intensity task. The procedure starts with the raw response and intensity level (Step 1), projects them into two dimensions (intensity levels by categorized expressions; Step 2), and applies the Bayesian update to recover the full response profile (Steps 3 and 4). This procedure is repeated for all expressions in both tasks independently for each participant.

signal: −.013 [−.0186, −.00674], $t(157) = -4.21$), disgust (intensity: −.015 [−.0217, −.00732], $t(157) = -3.99$; signal: −.017 [−.0249, −.00981], $t(157) = -4.55$), sadness (intensity: −.022 [−.0286, −.0151], $t(157) = -6.36$; signal: −.014 [−.0209, −.00614], $t(157) = -3.61$), and surprise (intensity: −.024 [−.0316, −.0164], $t(157) = -6.22$; signal: −.017 [−.0238, −.0096], $t(157) = -4.65$). Older participants were able to recognize an emotional expression with less information, at lower levels of signal or intensity, than younger participants.

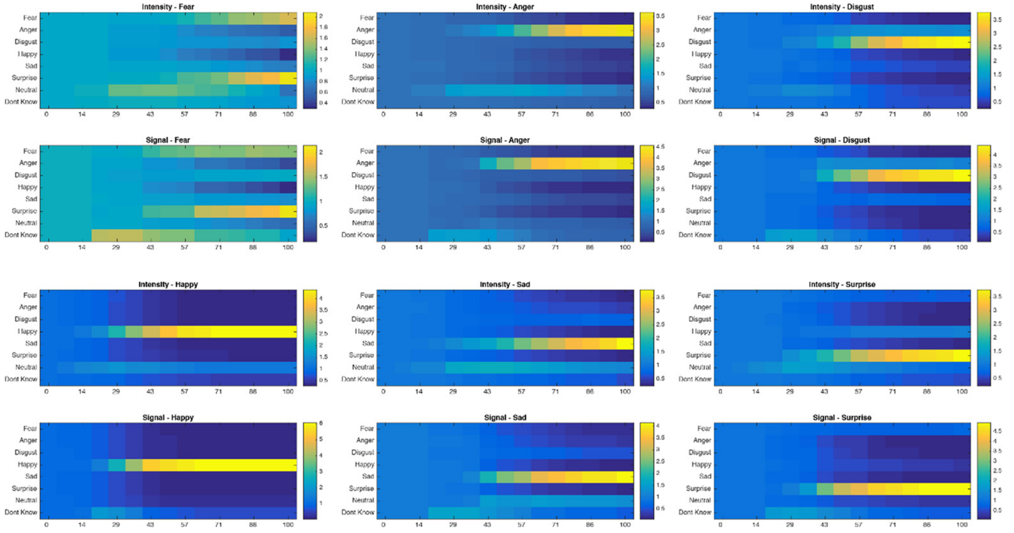

Fig. 6. Average response profiles across groups for the intensity and signal tasks and each facial expression of emotion. Each subplot shows the average response profile for one expression in one task. The rows represent the different responses from participants (six basic expressions + neutral + don't know response), whereas the columns represent the signal or intensity level presented by QUEST. The value of the color maps are the concentration parameter α of the Dirichlet distribution (see the main text and Fig. 5 for more details). A high value indicates higher probability *and* greater confidence in choosing that response. (For interpretation of the reference to color in this figure legend, the reader is referred to the Web version of this article.)

Response profile analysis

As shown in Fig. 6, there are substantial differences between the response profiles of the two tasks for most of the expressions. To establish how similar the signal and intensity measures were across development, we performed a mutual information analysis on the response profiles of both measures for each participant (Fig. 7). Each plotted point, therefore, represents the similarity in the response profiles of the two measures for one participant. Overall, there was an upward trend for the response profiles to become more similar with age, with four of the six emotions showing a significant increase in similarity with age: anger (regression coefficient: β = .014 [.0036, .0243], $t(157)$ = 2.67, p = .0083), disgust (β = .018 [.0073, .0286], $t(157)$ = 3.33, p = .0011), sadness (β = .012 [.0025, .0219], $t(157)$ = 2.48, p = .0143), and surprise (β = .024 [.0013, .0361], $t(157)$ = 4.06, p = 7.73e−5). Moreover, a robust GLM between mean mutual information across expression and age showed a significant positive correlation, β = .0103 [.006, .0145], $t(157)$ = 4.74, p = 4.75e−6. As the response profiles become more similar with age, erroneous responses become less random in comparison with younger participants.

Discussion

Using a psychophysical approach and an experimental design in which all participants completed both experimental conditions, we aimed to (a) isolate the quantities of *signal* and *intensity* (using neutral expression image morphs) necessary to recognize six prototypical facial expressions of emotion in children from 5 years of age up to adulthood while maintaining performance at 75% and (b) compare these measures to better understand the use and sensitivity of different measures in assessing recognition performance across development. To achieve these aims, we used a data-driven methodological approach by analyzing recognition performance on a *continuum* of age, a novel approach that overcomes the delimitation and use of arbitrary age boundaries.

The results of the first objective revealed that, as expected, the quantities of signal and intensity needed to recognize the majority of expressions decreased with age for sad, angry, disgust, and

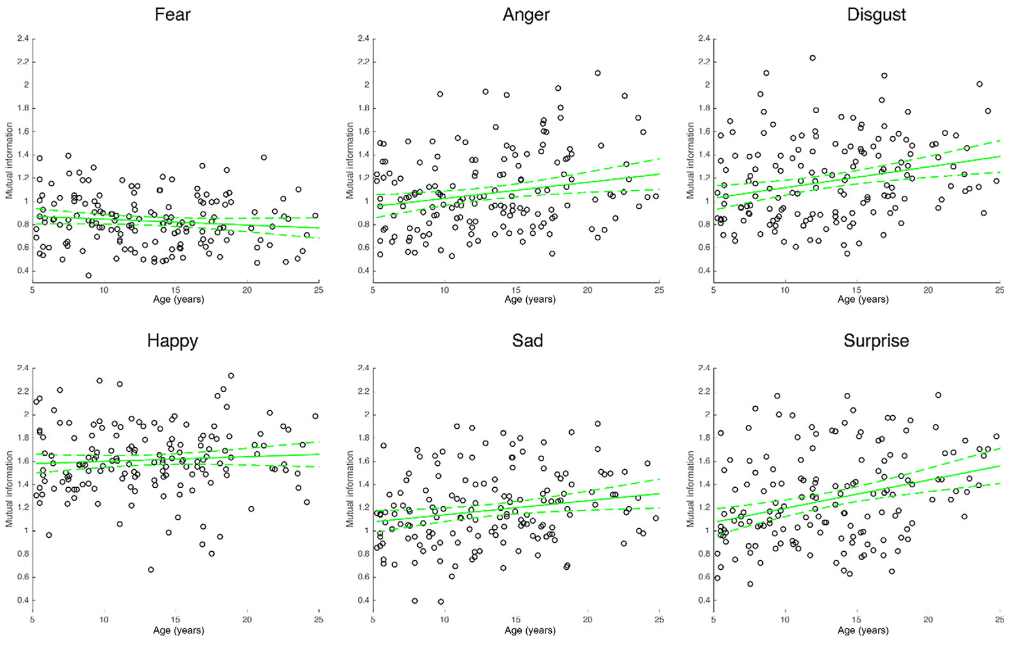

Fig. 7. Mutual information between the intensity and signal tasks for each participant per expression. Higher mutual information indicates higher similarity in the response profile between the two tasks. Line plots show the linear regression between mutual information and age, and dotted lines show the 95% confidence region of the regression.

surprise expressions, respectively. Therefore, the processing of both types of visual information becomes more discriminative during development as less information is needed with age to recognize these expressions. However, recognition improvement across development was not uniform for these expressions, as has also been shown in previous studies (Boyatzis, Chazan, & Ting, 1993; Gao & Maurer, 2010; Herba & Phillips, 2004; Lawrence, Campbell, & Skuse, 2015; Mancini et al., 2013; Rodger et al., 2015; Vicari et al., 2000). For fear and happy expressions, age did not have a major impact on the quantities of signal and intensity use. Therefore, recognition performance for fear and happy expressions was relatively stable from 5 years of age. For both measures, fear and happiness were the most difficult and easiest expressions to recognize across age groups, respectively. Earlier studies have similarly shown that happy expressions have the highest recognition performance and that this remains stable from an early age (Gao & Maurer, 2009, 2010; Gross & Ballif, 1991; Herba & Phillips, 2004; Mancini et al., 2013). However, one recent study showed that despite the youngest age group tested (6-year-olds) showing 92% recognition accuracy for happy expressions, there was a small but significant improvement in accuracy with age (Lawrence et al., 2015). Stability in accuracy levels for fear recognition from an early age was similarly found in an earlier study measuring signal recognition thresholds uniquely (Rodger et al., 2015). Overall, the recognition thresholds for both measures showed a similar trend in improvement or stability across expressions and in the ease and difficulty of happy and fear recognition, respectively.

Although similar developmental trajectories for recognition of these expressions have been revealed in other studies, the reasons behind such particular patterns of trajectories remain widely speculated. Studies that have tested different developmental cohorts have shown the effect that experience has on emotion recognition. For example, a cross-cultural eye-tracking study testing Caucasian and Asian infants recently revealed information sampling biases in infants as young as 7 months when they discriminate facial expressions of emotion (Geangu et al., 2016; Caldara, 2017). Early culture-specific experience, therefore, can affect which visual information we sample from the environment. The authors speculated that within the cultural environment, it is possible that parental practices most prevalently affect young infants. For example, Asian mothers have been found to be less emotionally expressive and to use more non-direct body contact compared with Western mothers (Kisilevsky et al., 1998), which could affect infants' attentional strategies toward the culturally specific, emotionally salient features of the face and body.

Similarly, by testing different developmental cohorts, children who have been exposed to physical abuse and those who have not, Pollak and colleagues' work has notably demonstrated the effect of emotional experience on emotion recognition (Pollak, Cicchetti, Hornung, & Reed, 2000; Pollak & Kistler, 2002; Pollak, Messner, Kistler, & Cohn, 2009; Pollak & Sinha, 2002). Children who had suffered physical abuse were consistently found to recognize anger more rapidly, or with fewer physical cues, than non-abused children. Because the children studied had similar sociodemographic and family backgrounds except for the experience of physical abuse, the explanation of this heightened sensitivity for anger recognition alone suggests that affective experience can influence perceptual representations of emotions.

As described above, developmental studies of facial expression recognition have repeatedly shown that happy is the easiest and earliest expression to be recognized. This facility with happiness could be partially explained in typically developing children by our frequent exposure to smiling faces during early childhood combined with the visual distinctiveness of happiness from other expressions (Calvo & Marrero, 2009; Kohler et al., 2004). By contrast, expressions of fear, although critical to our survival, are not commonly experienced frequently during daily life. Although experience alone might not account for poor recognition rates of fear in adult and developmental studies (Calder et al., 2003; Gross & Ballif, 1991; Herba & Phillips, 2004; Rapcsak et al., 2000; Widen, 2013), it is a possible contributory factor. The low rates of fear recognition from both signal and intensity measures in our study suggest that for optimal recognition additional information is required, perhaps from several modalities. Experiential factors affecting the recognition of sadness have also been shown in studies of depressed adults (Arteche et al., 2011; Gollan, McCloskey, Hoxha, & Coccaro, 2010; Gur et al., 1992; Kluczniok et al., 2016). Precisely because measuring an individual's prior experience of emotional expressions empirically is difficult, measuring different cohorts to inform how cultural and social experiences affect our capacity to recognize emotions is valuable. Future cross-cultural, clinical, and developmental studies with diverse cohorts could adopt the paradigm here to establish possible differences in sensitivity to signal and intensity information and to further determine whether any differences found are related to experiential factors.

Comparison of response profiles for signal and intensity measures

To establish how comparable the signal and intensity measures were across development, and to better understand the use and sensitivity of different measures in assessing recognition performance across development, we compared the measures using a novel data-driven analysis. We used mutual information analysis to establish how similar the response profiles of the signal and intensity measures were for each individual on a *continuum* of age. The analysis of age in years on a continuous scale rather than a categorical scale is a data-driven, non-biased approach that permits a finer level of analysis to provide a more precise picture of how the development of facial expression recognition unfolds. The mutual information analysis showed that the response profiles of the signal and intensity measures became more similar with age for the sad, angry, disgust, and surprise expressions. Again, for fear and happy expressions, no significant change across development was evident. Similarity in the response profiles of the sad, angry, disgust, and surprise expressions was evident only in the oldest participants. Therefore, the response profiles for emotional expressions of full intensity in the signal condition did not correspond with the profiles obtained from expressions of varying intensities in the morph condition throughout the majority of development.

The mutual information analysis, therefore, established that two types of stimuli commonly used in facial emotion processing studies (expressions at full intensity vs. expressions of varying intensities) cannot be straightforwardly compared during development. This critical point is another explanatory factor, along with differences in age groups, expressions, and tasks, for the differences in recognition trajectories found throughout the developmental literature described in the Introduction.

Importantly, as the response profiles for sad, angry, disgust, and surprise expressions became more similar with age, erroneous responses become less random. This suggests that representations of emotional expressions are more robust in the oldest participants tested given that they produced systematic confusions, for example, fear for surprise. Therefore, novel analysis of the overall response profiles for the expression recognition tasks, rather than the more standard practice of analyzing the final response values per se, revealed subtle but important changes in the sequence of responses along the continuum of age.

The presence of more robust expression representations in adulthood aligns with Widen and Russell's differentiation model of emotion (Widen, 2013; Widen & Russell, 2003, 2008) that emotion concepts are acquired gradually throughout development, beginning with a broad concept including any emotion of the same valence, and hence the potential for greater confusion, with concepts gradually narrowing and becoming more discrete with age. Only visual information was available in the current study without any social or contextual information to aid accurate categorization. Thus, it is plausible that the randomness in the categorization errors of younger children might arise from their lack of sufficiently robust visual representations of an emotion despite already having a concept of the emotion. A potential mechanism for this refining of emotion categories and greater robustness in their perceptual representations is provided by Leppänen and Nelson (2009). In their review of how the developing brain becomes tuned to the social signals of emotional expressions, they described an experience-dependent mechanism that is necessary for the development of a mature system. They proposed that our perceptual representations of facial expressions are initially coarsely specified and develop into a mature system with adult-like specificity only through exposure to species-typical emotional expressions. The experience-dependent nature of facial expression processing has been shown by the disruption caused to typical development by species-atypical parenting and social deprivation (e.g., Pollak & Kistler, 2002). In contrast, typical development, as shown here, results in a mature system with more highly specified categorical representations of expressions that are also prone to more systematic errors. Future studies can apply these methods to further investigate sensitivity to specific emotions and discrepancies in response confusions across the lifespan as well as in diverse clinical groups.

Future studies should also investigate some of the limitations of the current study. This cross-sectional study has revealed specific developmental trajectories and response profiles for expression recognition using signal and intensity information. A longitudinal design with neural measures could further establish how processing of signal and intensity information evolves with age, for example, whether neural populations processing the two types of information overlap with age as sensitivity for decoding both types of information becomes more similar. It is also worth noting that there is large variability for the mutual information estimation between the response profiles from the two tasks. Although our data show a significant association between age and the estimated mutual information, it is important to consider the practical significance of the observed effect. Indeed, the coefficient estimation values from the regression model are generally between .01 and .025, which is not a large change compared with the intercept (with an estimate of roughly 1 bit). However, no studies have compared the change of multivariate response patterns in a behavioral task using mutual information; thus, it is difficult to straightforwardly interpret the current changes of information use. Further studies are necessary to quantify the practical significance of the effects observed here. Similarly, adult face stimuli uniquely were studied. The question of whether there is an own-age advantage for emotion recognition is still debated (Griffiths, Penton-Voak, Jarrold, & Munafò, 2015; Hills, 2012; Wiese, Komes, & Schweinberger, 2013). It is possible that with own-age stimuli, children may recognize emotional expressions alternatively and show different developmental trajectories for recognition than is found with adult faces. Finally, because faces are not the only signal used in the natural environment for effective emotional communication, future studies are needed to further evaluate how contextual and other social cues contribute to the processing of emotional expressions across development to determine, for example, the effect of non-facial emotional cues on recognition when face signals or intensity is modified as it is here.

Conclusions

These findings have important theoretical and methodological implications for developmental, lifespan, emotion, and face processing research. First, findings from facial expression recognition studies with different age cohorts using emotional expressions of full intensity, as in the signal condition here, cannot be straightforwardly compared with findings using varied expression intensities. Throughout development, the response profiles for recognition of expressions at full intensity were not comparable to those of varied intensities. Second, by examining individual responses along an age continuum, as opposed to the mean responses of age group categories, a finer level of analysis is possible that can provide a more precise picture of differences occurring during development. Here, this revealed a gradual reduction in information use for recognition of four of the six expressions tested and, for the same expressions, a gradual increase in the similarity of response profiles with age. Third, by analyzing the response profiles (i.e., the sequence of responses across trials) rather than the fixed end-point measure of recognition score, a richer explanation of what is occurring is possible as we compare the overall distribution of responses. For example, this approach revealed that the response profiles become more similar with age due to less random erroneous categorizations. This broader analysis, therefore, can provide insight into the underlying processing of visual information; because the categorization errors become less random with age, this suggests that the expression representation becomes more robust. Potentially, therefore, the neural populations processing the two types of information—signal and intensity—overlap with age as sensitivity for decoding both types of information becomes more similar. Altogether, our data provide novel methodological and theoretical insights into the developing affective system.

Acknowledgments

We especially thank all of the participants from various schools in the canton of Fribourg, Switzerland, for being willing to take part in our study: Ecole Enfantine et Primaire de Marly Grand Pré, Cycle d'Orientation de Pérolles, and Ecole de Culture Générale Fribourg. We also thank all of the teachers for their patience with classroom interruptions as well as head teacher Dominique Brulhart, principal teacher Claude Fragnière, and technician Christophe Pochon for their essential organization of our visits. Furthermore, we thank Bérangère Prongue, Elena Bayboerek, Sonia Pupillo, Vincent Junod, and Faustine Etter for their help with data collection.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jecp.2018.05.005.

References

Arteche, A., Joormann, J., Harvey, A., Craske, M., Gotlib, I. H., Lehtonen, A., & Stein, A. (2011). The effects of postnatal maternal depression and anxiety on the processing of infant faces. *Journal of Affective Disorders, 133*, 197–203.
Boyatzis, C. J., Chazan, E., & Ting, C. Z. (1993). Preschool children's decoding of facial emotions. *Journal of Genetic Psychology, 154*, 375–382.
Bruce, V., Campbell, R. N., Doherty-Sneddon, G., Langton, S., McAuley, S., & Wright, R. (2000). Testing face processing skills in children. *British Journal of Developmental Psychology, 18*, 319–333.
Caldara, R. (2017). Culture reveals a flexible system for face processing. *Current Directions in Psychological Science, 26*, 249–255.
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.
Calvo, M. G., & Marrero, H. (2009). Visual search of emotional faces: The role of affective content and featural distinctiveness. *Cognition & Emotion, 23*, 782–806.
Carton, J. S., Kessler, E. A., & Pape, C. L. (1999). Nonverbal decoding skills and relationship well-being in adults. *Journal of Nonverbal Behavior, 23*, 91–100.
De Sonneville, L. M. J., Vershoor, C. A., Njiokiktjien, C., Veld, V. O. H., Toorenaar, N., & Vranken, M. (2002). Facial identity and facial emotions: Speed, accuracy, and processing strategies in children and adults. *Journal of Clinical and Experimental Neuropsychology, 24*, 200–213.
Feldman, R. S., Philippot, P., & Custrini, R. J. (1991). Social competence and nonverbal behavior. In R. S. Feldman & B. Rime (Eds.), *Fundamentals of nonverbal behavior* (pp. 329–350). New York: Cambridge University Press.

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.
Geangu, E., Ichikawa, H., Lao, J., Kanazawa, S., Yamaguchi, M. K., Caldara, R., & Turati, C. (2016). Culture shapes 7 month olds' perceptual strategies in discriminating facial expressions of emotion. *Current Biology, 26*, R663–R664.
Gollan, J. K., McCloskey, M., Hoxha, D., & Coccaro, E. F. (2010). How do depressed and healthy adults interpret nuanced facial expressions? *Journal of Abnormal Psychology, 119*, 804–810.
Gosselin, F., & Schyns, P. G. (2001). Bubbles: A technique to reveal the use of information in recognition tasks. *Vision Research, 41*, 2261–2271.
Griffiths, S., Penton-Voak, I. S., Jarrold, C., & Munafò, M. R. (2015). No own-age advantage in children's recognition of emotion on prototypical faces of different ages. *PLoS One, 10*(5). e125256.
Gross, A. L., & Ballif, B. (1991). Children's understanding of emotion from facial expressions and situations: A review. *Developmental Review, 11*, 368–398.
Gur, R. C., Erwin, R. J., Gur, R. E., Zwil, A. S., Heimberg, C., & Kraemer, H. C. (1992). Facial emotion discrimination: II. Behavioral findings in depression. *Psychiatry Research, 42*, 241–251.
Hall, E. (1966). *The hidden dimension*. Garden City, NY: Doubleday.
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*, 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.
Hills, P. J. (2012). A developmental study of the own-age face recognition bias in children. *Developmental Psychology, 48*, 499–508.
Izard, C., Fine, S., Schultz, D., Mostow, A., Ackerman, B., & Youngstrom, E. (2001). Emotion knowledge as a predictor of social behavior and academic competence in children at risk. *Psychological Science, 12*, 18–23.
Johnston, P. J., Kaufman, J., Bajic, J., Sercombe, A., Michie, P. T., & Karayanidis, F. (2011). Facial emotion and identity processing development in 5- to 15-year-old children. *Frontiers in Psychology, 2*. https://doi.org/10.3389/fpsyg.2011.00026.
King-Smith, P. E., Grigsby, S. S., Vingrys, A. J., Benes, S. C., & Supowit, A. (1994). Efficient and unbiased modifications of the QUEST threshold method: Theory, simulations, experimental evaluation, and practical implementation. *Vision Research, 34*, 885–912.
Kinney, J. B., & Atwal, G. S. (2014). Equitability, mutual information, and the maximal information coefficient. *Proceedings of the National Academy of Sciences of the United States of America, 111*, 3354–3359.
Kisilevsky, B. S., Hains, S. M., Lee, K., Muir, D. W., Xu, F., Fu, G., ... Yang, R. L. (1998). The still-face effect in Chinese and Canadian 3- to 6-month-old infants. *Developmental Psychology, 34*, 629–639.
Kluczniok, D., Attar, C. H., Fydrich, T., Fuehrer, D., Jaite, C., Domes, G., ... Bermpohl, F. (2016). Transgenerational effects of maternal depression on affect recognition in children. *Journal of Affective Disorders, 189*, 233–239.
Kohler, C. G., Turner, T., Stolar, N. M., Bilker, W. B., Brensinger, C. M., Gur, R. E., & Gur, R. C. (2004). Differences in facial expressions of four universal emotions. *Psychiatry Research, 128*, 235–244.
Lawrence, K., Campbell, R., & Skuse, D. (2015). Age, gender, and puberty influence the development of facial emotion recognition. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.00761.
Leppänen, 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*. Stockholm, Sweden: Karolinska Institute.
Mancini, G., Agnoli, S., Baldaro, B., Ricci Bitti, P. E., & Surcinelli, P. (2013). Facial expressions of emotions: Recognition accuracy and affective reactions during late childhood. *Journal of Psychology: Interdisciplinary and Applied, 147*, 599–617.
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.
Nowicki, S., Jr., & Duke, M. P. (1992). The association of children's nonverbal decoding abilities with their popularity, locus of control, and academic achievement. *Journal of Genetic Psychology, 153*, 385–394.
Pelli, D. G. (1987). The ideal psychometric procedure. *Investigative Ophthalmology & Visual Science, 28*, 366.
Pollak, S. D., Cicchetti, D., Hornung, K., & Reed, A. (2000). Recognizing emotion in faces: Developmental effects of child abuse and neglect. *Developmental Psychology, 36*, 679–688.
Pollak, S. D., & Kistler, D. (2002). Early experience alters categorical representations for facial expressions of emotion. *Proceedings of the National Academy of Sciences of the United States of America, 99*, 9072–9076.
Pollak, S. D., Messner, M., Kistler, D. J., & Cohn, J. F. (2009). Development of perceptual expertise in emotion recognition. *Cognition, 110*, 242–247.
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., Kaszniak, A. W., ... Cohen, R. A. (2000). Fear recognition deficits after focal brain damage. *Neurology, 54*, 575–581.
Rodger, H., Vizioli, L., Ouyang, X., & Caldara, R. (2015). Mapping the development of facial expression recognition. *Developmental Science, 18*, 926–939.
Schyns, P. G., Bonnar, L., & Gosselin, F. (2002). Show me the features! Understanding recognition from the use of visual information. *Psychological Science, 13*, 402–409.
Smith, M. L., Gosselin, F., Cottrell, G. W., & Schyns, P. G. (2005). Transmitting and decoding facial expressions. *Psychological Science, 16*, 753–761.
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*, 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.

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*, 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.

Widen, S. C., & Russell, J. A. (2003). A closer look at preschoolers' freely produced labels for facial expressions. *Developmental Psychology, 39*, 114–128.

Widen, S. C., & Russell, J. A. (2008). Children acquire emotion categories gradually. *Cognitive Development, 23*, 291–312.

Wiese, H., Komes, J., & Schweinberger, S. R. (2013). Ageing faces in ageing minds: A review on the own-age bias in face recognition. *Visual Cognition, 21*, 1337–1363.

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