Seven- to 11-year-olds' developing ability to recognize natural facial expressions of basic emotions

Kathleen Kang, Laura Anthoney, Peter Mitchell

School of Psychology, University of Nottingham,

University Park, Nottingham NG7 2RD, UK

Tel 0115 9515313

peter.mitchell@nottingam.ac.uk

Running head: Recognizing natural expressions

<u>Abstract</u>

Being able to recognize facial expressions of basic emotions is of great importance to social development. However, we still know surprisingly little about children's developing ability to interpret emotions that are expressed dynamically, naturally and subtly, despite real-life expressions having such appearance in the vast majority of cases. The current research employs a new technique of capturing dynamic, subtly expressed natural emotional displays (happy, sad, angry, shocked and disgusted). Children aged 7, 9 and 11 years (and adults) were systematically able to discriminate each emotional display from alternatives in a 5-way choice. Children were most accurate in identifying the expression of happiness and were also relatively accurate in identifying the expression of sadness; they were far less accurate than adults in identifying *shocked* and *disgusted*. Children who performed well academically also tended to be the most accurate in recognizing expressions and this relationship maintained independently of chronological age. Generally, the findings testify to a well-developed ability to recognize very subtle naturally occurring expressions of emotions.

Seven- to 11-year-olds' developing ability to recognize natural facial

expressions of basic emotions

Being able to recognize facial expressions of emotion is critical to social development (Balaban, 1995), is central to the emerging ability to understand other minds (Baron-Cohen, Joliffe, Mortimore & Robertson, 1997) and more generally plays a role in survival in the broadest sense (Ekman, 1992). Not surprisingly, then, the development of this important ability has been widely researched over many years (McClure, 2000). The published literature tells us that the earliest signs in recognizing and classifying emotional expressions are evident in the first year of life (see Lappänen & Nelson 2009, for a review), with the earliest recognition of some expressions apparent in neonates (Farroni, Menon, Rigato & Johnson, 2007). Performance in recognizing such expressions improves over childhood (Herba & Phillips, 2004) and females might develop more rapidly than males (Hall, 1984), though this gender difference seems to disappear when emotional expressions are presented in a more naturalistic context (Theurel et al, 2016).

Arguably, the accumulated knowledge summarised above sometimes is limited by over-reliance on rather unnatural stimuli -- standardised static photographs of exaggerated posed expressions (McClure, 2000; Widen & Russell, 2015). In real life we are normally faced with the task of interpreting very subtle and perhaps barely visible expressions (Rodger, Vizioli, Ouyang & Caldara (2015) and it is thus legitimate to enquire how development progresses when the signal is not strong, as in real life. To investigate this matter, Rodger et al presented systematically degraded images of facial expressions in order to determine developmental changes in detection threshold. Using this technique, the researchers discovered gradually improved detection of emotional expressions of *sad*, *surprised*, *disgust* and *anger*, with a slightly steeper developmental trend in recognizing the latter two compared with the former two. There was no developmental trend associated with *happy* and *fear* in that *happy* was recognized easily by all age–groups while *fear* was difficult for all age

groups. The stimuli used by Rodger et al were actually strong expressions, giving a strong signal, but ones which were difficult to perceive on account of filters being used that meant it was as if participants were looking through fog.

Under what circumstance might people (targets) express emotions subtly such that the signal is expressed weakly, as is commonly the case in real life? Subtle expressions can occur when feelings are not intense or when targets mask strong feelings (Carroll & Russell, 1997). Relevant to the current research, a further example is when the target is engaged in social interaction and he or she reflects the expression worn by a third party (Chartrand and Bargh, 1999), indicating social resonance and connectedness (Stel and Vonk, 2010). If an observer could nevertheless accurately interpret the target's expression, then not only would they know something of the target's inner state, they would also effectively have access to a broader social world through the lens of the target's mind (Teoh et al, 2017). If so, the observer is parasitic on another mind (the mind of the target) to know something more broadly about the social world (that the person the target is speaking with is happy, angry, disgusted, etc); something they would not have known had they not been able to interpret the target's expression (Pillai et al, 2014).

The question then becomes, 'how does development unfold in the ability to recognize subtle, dynamic emotional expressions, of the kind one might encounter in real life?' The study by Rodger et al (2015) makes an important contribution by raising this question; the purpose of the current research is to advance knowledge further by tackling the extremely difficult problem of creating stimuli that offer but a weak signal of the emotion being expressed by the target. Rodger et al addressed this by systematically degrading the image (a facial expression) in order to determine the threshold at which children of various ages could detect the expression. Such a manipulation might be relevant to real life circumstances in which the target is wearing a very strong expression but the viewing conditions are less than

ideal, perhaps caused by ambient fog or other conditions of poor visibility. However, in real life people seldom wear strong facial expressions. Rather, real life expressions are subtle, denoted by minute changes in one or just a few discrete facial features (Carroll and Russell, 1997). Consequently, the challenge normally faced by the observer is to detect an expression that is difficult to interpret not because of poor viewing conditions but because the target's expression is naturally and inherently subtle, making the signal weak.

With the above in mind, it was high priority that the facial expressions were subtle, natural, spontaneous and dynamic (not static photographs – cf Widen & Russell, 2015). We achieved this by showing young adults (the targets) a series of intense facial expressions on a laptop computer screen. As a cover activity, targets were asked to select a cartoon expression of the emotion (i.e. emoticon) to match the displayed expression, as if the purpose of the task were to determine how well they could recognise the expressions. The true purpose of the task, though, was to surreptitiously record the target´s expression by the laptop's webcam (i.e., imitating expressions; Dimberg et al., 2000; Sato & Yoshikawa, 2007). Subsequently, we showed these videos of targets to "perceivers" (children and adults) to determine if they could identify the target's reflected emotional expression. Research suggests a gradual improvement over childhood in the detection of expressions (Rodger et al, 2015; Widen & Russell, 2015) and accordingly we predicted the same here. Research also suggests that *happy* is easier to detect than other expressions (McClure, 2000) but *disgust* can be difficult to detect (Widen & Russell, 2008), and this is also something we expected to find. More generally, some research suggests that females might be more accurate than males in recognizing emotional expressions (e.g. Montagne, Kessels, Frigerio, de Haan & Perrett, 2005), though evidence has not always been supportive of a gender difference (McClure, 2000; Theurel et al, 2016).

Accurate labelling of facial expressions is associated with social competence (Barth & Bastiani, 1997) and is also associated with burgeoning verbal skills which in turn promotes academic ability (Izard et al, 2001). In the current study, we recorded children's academic grades to determine if their detection of subtle expressions is indeed associated with this variable. Children's academic grade duly increases with age and therefore it will be important to establish that any relationship between emotion expression recognition and academic grade is not dependent on concordant increases in chronological age.

Method

Participants

One hundred and nine children were recruited from a mainstream Nottinghamshire primary school. One female was later excluded because her academic grades were not available. The children were in school-year 2 (18 males, 18 females; mean age, 7.5 years; range 7.0-8.1), 4 (18 males, 21 females; mean age, 9.4 years; range, 9.0-9.9) and 6 (16 males, 18 females; mean age, 11.4 years; range 11.0-12.0) and were of mixed academic ability. These groups will henceforth be labelled as 7-year-olds, 9-year-olds and 11-year-olds. All were British nationals and native English speakers. Pilot work suggested that children below 7 years had difficulty working with the computer application developed for this study (see below).

Permission was obtained from the headmaster for testing to be conducted in school. An information sheet and consent letter was distributed to all parents/guardians of children in the selected school years. The letter contained general information about the study along with an explanation that they were under no obligation to have their child participate. Each child was asked individually by the researcher if they wished to participate/withdraw; all whose parents had given consent elected to participate. All of the participants received a small incentive of either stickers or a choice of a small stationary item (e.g. eraser/pencil).

Apparatus, materials & stimuli

The target stimuli were 3-second video clips captured by the webcam of a laptop, of young adults (targets) as they looked at a succession of facial expressions displayed on a Toshiba Satellite CS650-S5501 laptop with a 15.6 inch screen. The expressions viewed by the targets were taken from the KDEF database (Karolinska Directed Emotional Faces -- Lundqvist et al. 1998) which consisted of 10 actors (5 males, 5 females) displaying happy, sad, angry, surprised (shocked), disgusted and neutral expressions selected at random using MS Excel 2007. Fear was not included as it is often classified less accurately (Calvo and Lundqvist, 2008) due to confusion with angry or disgusted expressions (Jack et al., 2009). The size of these photographs was 562 x 762 pixels and the resolution was 72 x 72 dpi. While each target viewed an emotional expression they were tasked with selecting an emoticon that matched the expression (Figure 1) and as they did so, unknown to them, they were recorded by the laptop's integral webcam. As they viewed each facial expression, targets spontaneously tended to faintly imitate what they saw (Figure 2). Targets also always chose the appropriate emoticon which corresponded to the expression that they viewed. For instance, when targets were viewing a happy expression, they always chose the happy emoticon.

Video recordings of targets' expressions were then edited to produce individual video clips which represented the targets' reactions to a particular expression that they were viewing. As there were 30 targets (15 males, 15 females, Mean age = 20.27 years, SD = 2.36 years) who viewed 60 emotional expressions each, the total number of edited video clips was 1800. As targets always made their responses within 3 seconds, targets' expressions were trimmed to capture the first three seconds as they viewed each face stimulus. From the 1800 individual clips, 77 were selected on the basis that there was visible facial activity in the target. The videos were independently rated by the first author and another person and the

inter-rater agreement was *Kappa* = 0.83 (*p*<0.001). The remaining 1723 individual clips were entered into MS Excel and 163 videos were selected at random. Thus, a total of 240 video clips (77 + 163) were used in the validation stage. Ten judges (5 females, 5 males) independently assessed the expressiveness of the video clips from a scale of 1 (least expressive) to 6 (most expressive). Based on the ratings of expressiveness, out of these 240 clips, 10 were selected as the most expressive imitations of each emotional expression. Targets' reactions to neutral faces were omitted resulting in 50 video clips for presentation to the children.

- - - - - - - - - - - - - - - -

Figure 1 about here

- - - - - - - - - - - - - - - -

The task was presented to children using PsychoPy Software v1.82.02 (Peirce, 2007). An instruction slide appeared and after reading this (with the experimenter's help) the participants could press any key to move on to the practice trial. All video clips, including the practice, were displayed for three seconds and were positioned centrally in the screen at a size of 16 (W) x 14 (H) cm. At the bottom of each screen the five emoticons were displayed at a size of 2 (W) x 2 (H) cm for each emoticon (see Figure 1). Each video clip started automatically after the participant had selected their chosen emoticon for the previous video. Upon completion of all fifty trials, a screen displaying 'thank you for participating' notified the participants that they had completed the task. Videos were presented in a random order by the software to each participant.

The task was presented to children with a Sony Vaio laptop (16inch) and the participants sat approximately 60cm from the screen. Each participant was given the choice of using either the keypad or the hand held mouse to select an emoticon for each target video.

The five emoticons were also printed onto A4 sheets for the purpose of establishing how children defined each of these prior to testing.

Procedure

Each child participated in the task individually and without time constraint. They began by looking at the five emoticons printed on a sheet of A4 and gave a description of how each was feeling (happy, sad, angry, shocked or disgusted). Subsequently, participants viewed an instruction slide on the laptop ("Which emoticon did the person choose?"), which the experimenter read aloud as they did so. The slide explained that targets had looked at photographs of faces showing a particular expression and had then chosen an emoticon to match that expression. Child participants were asked to guess which emoticon the target had chosen, out of the five available. Note that the question could simply ask how each target was feeling but the relationship between a facial expression and an inner state is rather imponderable such that the accuracy of any response is difficult to determine. In the interest of transparency, children were thus asked to infer the target's choice of emoticon, allowing unequivocal comparison between their response and an objective fact.

Participants were shown a practice video clip and asked to select their chosen answer out of the five emoticons displayed at the bottom of the screen. They could make their choice either by selecting a number on the keypad (1-5) or by clicking one of the emoticons using the mouse, as they preferred. Prior to moving on to the actual task they were again asked whether they wished to participate/withdraw from the study and all gave their assent. Note that in the target videos (Figure 2) there was absolutely no clue to which emoticon targets had chosen other than the target's spontaneous and faint imitation of the expression they were viewing.

- - - - - - - - - - - - - - - -

Figure 2 about here

- - - - - - - - - - - - - - - -

Data analysis

PsychoPy software produced an output data file which was converted into Microsoft Excel 2010 for each participant, containing their scores and response times from the task. From these files, five matrices per participant were derived (one matrix for each emotion, happy, sad, angry, shocked and disgusted), detailing the number of hits (e.g. choosing *happy* when the target had indeed selected a happy emoticon), false alarms (e.g. choosing *happy* when the target had selected something other than a happy emoticon), misses (e.g. choosing *sad* when the target had actually selected a happy emoticon) and correct rejections (e.g. choosing something other than *happy* when the target had indeed selected something other than a happy emoticon). These matrices were used to calculate five *d*-prime values for each participant which resulted in 540 (5 emotions X 108 participants) *d*-prime values in total, forming a normally distributed set of data that are suitable for analysis using parametric statistical tests (Pillai et al, 2012, 2014). The *d*-prime values were automatically calculated using an Excel algorithm (available from the authors on request) and gave an indication of successful performance, corrected for any bias in choosing particular response options, by comparing the normalised scale of hits against normalised false alarms for each participant for each kind of target expression.

Ethical approval was obtained from the School of Psychology, University of Nottingham, independently for the target phase and participant phase of the study. Due to the participants of the present study being under the age of 18 years, informed consent was obtained from the parents/guardians prior to any child giving their own assent.

Results

Children in all age groups were able to give accurate verbal labels for the emoticons happy, sad, angry and shocked. In contrast, only 11% of 7-year-olds, 40% of 9-year-olds and 20% of 11-year-olds labelled the disgust emoticon appropriately, suggesting that this word might not be salient in their vocabulary (Widen & Russell, 2008). Despite not knowing the word *disgust*, this should not necessarily prevent children from judging that they thought the target had chosen this emoticon after viewing the target's facial expression. In other words, successful performance in the main task did not necessarily depend upon being able to label the emoticons.

Table 1 is a confusion matrix comparing emoticons chosen by targets with emoticons chosen by child participants. The values highlighted in grey represent 'hits' -- in other words, occasions when children chose the same emoticon that targets chose. As the table shows, hits were more common than any other kind of response but with two exceptions (7-year-olds most commonly chose a 'shocked' emoticon when observing a target who was actually viewing an angry expression and 9-year-olds most commonly chose a 'disgusted' emoticon when observing a target who was actually viewing a shocked expression).

- - - - - - - - - - - - - - - -

Table 1 about here

- - - - - - - - - - - - - - - -

d-prime values were calculated as described above, where any mean value of 0 would indicate that the false alarm and hit rates were identical, suggesting inability to discriminate which emoticon targets had chosen. Figure 3 shows 20 mean *d*-prime values, representing three age groups of children (plus adults – see below) each making judgments about 5 emotions. The figure shows that all means are greater than zero, and this was confirmed by one-sample *t*-tests ($p \leq .001$ in all cases), demonstrating that participants were systematically

able to determine which emoticon targets had chosen as they (the targets) looked at photos of 5 different emotional expressions.

- - - - - - - - - - - - - - - -

Figure 3 about here

- - - - - - - - - - - - - - - -

A 2 (gender) x 3 (age of child) x 5 (emoticon chosen by target) mixed Analysis of Variance (ANOVA), the last factor being a repeated measure, was conducted on these mean d-prime values. Partial eta squared (ηp^2) is reported to give an indication of effect sizes. A large (Cohen, 1992) main effect was found for the particular emoticon chosen by targets, $F(4, 408) = 72.63$, $p < .001$, $\eta p^2 = .41$ and there was a small main effect associated with age, $F(2, 102) = 3.09$, $p = .052$, $\eta p^2 = .06$. All other effects were nonsignificant, including the main effect associated with gender, $F<1$. An analysis of simple effects associated with age showed that the mean score for 7-year-olds ($M = .55$, $SD = .33$) was significantly less than that of 11-year-olds ($M = .75$, $SD = .34$). However, the mean for 9-year-olds ($M = .63$, $SD = .30$) did not significantly differ from the other two age-groups. Generally, these results suggest that children performed systematically in judging which emoticon was chosen by the targets in the video clips. They also suggest a small performance increase with age.

To interpret the main effect associated with the emoticon chosen by the target, data were combined over the three age groups and then ten post hoc paired samples t-tests were conducted to identify when participants were most accurate in inferring which emoticons the targets had chosen. Bonferroni correction ($p = .005$) was used to minimise the possibility of finding false positive results. The happy emoticon was significantly more accurately selected than any of the other emoticons; sad ($t(107) = 10.05$, $p < .001$), angry ($t(107) = 15.75$, $p < .001$), shocked ($t(107) = 15.39$, $p < .001$) and disgusted ($t(107) = 11.51$, $p < .001$). The sad emoticon was significantly more accurately selected than the angry ($t(107) = 4.17$, $p < .001$)

and shocked ($t(107)$ =3.70, $p < .001$) emoticons. Other emoticons were not significantly more accurately selected than each other.

Three academic achievement grades for each child (reading, numeracy and writing) were provided as numerical values which corresponded with their national curriculum grade. A Pearson's correlation revealed that these values highly intercorrelated for each year group independently, ranging from $r=.95$ to $r=.75$. It thus seemed appropriate to calculate a mean academic grade for each child, based on their three scores: 7-year-olds, mean 15.84, SD 2.11, range 8.30-18.70; 9-year-olds, mean 23.50, SD 2.47, range 15.30-27.00; 11-year-olds, mean 30.73, SD 3.06, range 24.00-35.30. These scores are typical for the age groups according to the UK government's data on National Curriculum Key Stage 2 (https://www.gov.uk/national-curriculum/overview).

A Pearson's correlation was then conducted to investigate the relationship between the children's performance on the task (average d-prime scores) and average academic grade. A positive correlation was found, $r = .31$, $p = .001$, indicating that children who were performing at a higher academic level were also performing well in guessing which emoticon the target had chosen. Not surprisingly, older children tended to have higher academic grades than younger children. However, the significant correlation between average d-prime scores and academic grade was not entirely explained by both of these variables correlating also with chronological age: After partialling out age, the correlation between academic grade and participants' average d-prime score remained significant, $r(105) = .29$, $p=.002$. Hence, those children who were performing well academically were also the most effective in determining which emoticon targets had chosen (and therefore which expression targets were looking at), and this was not just because academically more able children also happened to be older.

The total time each child took to complete the task was recorded -- the means (and standard deviations) for each year group were: 7-year-old mean = 285.09 s, ($SD = 74.57$), 9-

year-old mean = 284.36 s, (*SD* = 76.27) and 11-year-old mean = 276.50 s, (*SD* = 65.81). A one-way ANOVA offered no evidence to suggest that time taken to complete the task varied between the year groups, *F*<1. Moreover, there was no evidence to suggest that the time taken to complete the task was related with their average *d*-prime scores (an index of how accurately they detected which emoticon targets had chosen), *r* = .12, *p* = .22. Hence, there was no evidence of a "speed-accuracy trade-off" and there is no reason to suppose that children who performed well on the task did so because they occupied more time in finding the correct answer.

Adult comparison

While children in all age groups were significantly accurate in inferring which emoticon the target had selected, it is legitimate to ask why their level of accuracy was not greater. Was it because the task is difficult (difficult even for adults) or was it because the participants, being children, were not fully developed? If the former, we would expect to find accuracy at a similar level in adult participants; if the latter, then we would expect to find rather higher levels of accuracy in adults.

As with the procedure for children, 26 participants aged between 18 to 30 years old (13 females, 13 males) were told that they will be viewing targets' reactions to various expressions, though the task was scaled up for adults. They were tasked with selecting the emoticon they thought targets had chosen and responded by clicking on the emoticon from 6 alternatives (happy, sad, angry, shocked, disgusted, neutral). There were 20 trials for each type of expression, amounting to 120 in total. The task was presented with PsychoPy v.1.83.

In order to compare between children's and adults' performance, the *d*-prime values specifically for the 50 trials used in the child study were extracted from the adult data (Figure 3). Participants were significantly accurate for all expressions [happy, $t(25)=12.49$, $p<0.001$; sad, $t(25)=4.73$, $p<0.001$; angry, $t(25)=2.42$, $p=0.02$; shocked, $t(25)=7.71$, $p<0.001$;

disgusted, $t(25)=10.41$, $p<0.001$]. Because the purpose of the adult comparison was specifically to compare levels of accuracy between adults and children, we shall focus on the main effect of age and any interactions with age. We thus repeated the focal ANOVA reported above, except with four rather than three levels of age: 2 (gender) x 4 (year group) x 5 (emoticon chosen by target), the last factor being a repeated measure. There was a main effect associated with participant age, $F(3,126)=7.47$, $p<0.001$, $\eta p^2=0.15$, and participant age interacted with the emoticon chosen by targets, $F(12,504)=5.05$, $p<0.001$, $\eta p^2=0.11$.

In order to explain the interaction four separate one-way ANOVAs with four age levels were conducted for each kind of emoticon selected by targets. For the happy emoticon, there was a significant main effect of age, $F(3,133)=2.85$, $p=0.04$. Pairwise comparisons with LSD correction showed that adults performed significantly more accurately than 7-year-olds and 9-year-olds. There was also a significant main effect of age for shocked, $F(3,133)=7.93$, $p<0.001$, and for disgusted, $F(3,133)=13.00$, $p<0.001$. Adults performed significantly more accurately than all three child groups for both emotions. In summary, adults were more accurate than any of the child groups for shocked and disgusted but there was no evidence to suggest they were any more accurate for sad and angry.

Discussion

The results strongly show that children in all three age groups (and adults) for all five subtley expressed emotions were systematically accurate in determining which emoticon the target had selected. As predicted, there was a gradual age trend in that older children were more accurate in determining which emoticon targets had selected, a finding that is consistent with previous research (Rodger et al, 2015). Children were less accurate than adult participants when targets had selected shocked and disgusted emoticons; seven- and 9-year-old children were less accurate than adults when targets had selected a happy emoticon but not when they had selected sad and angry emoticons. Generally, participants were most

accurate in detecting when the target was viewing a happy expression, again as predicted, though the finding that children were better at detecting when the target was viewing a sad expression, compared with angry, shocked and disgusted, was not entirely consistent with previous research (Rodger et al, 2015). Neither was there any evidence to suggest that females outperformed males in recognizing emotional expressions (*pace*, Montagne et al, 2005), though this might be explained by contextualized presentation of expressions in the current study (Theurel et al, 2016).

As predicted, children's accuracy in determining which expressions targets were viewing correlated with their academic grade, and this was consistent with previous research (Izard et al, 2001). The higher levels of accuracy observed in children with higher grades was not entirely explained by their chronological age; even when the latter was partialled out, the relationship between accuracy and academic grade remained significant. While it is informative to know that being able to interpret emotional expressions is related with academic grade, we do not yet know the direction of causality. It might be that an individual who is good at interpreting emotional expressions also enjoys a level of social control that fosters their involvement in school work and bolsters their esteem such that they perform well academically (Trantacosta & Izard, 2007).

The method developed for this research has some strengths and also some limitations. Starting with the strengths, the task offers an effective way of capturing spontaneous and natural expressions that are subtle rather than extreme, as one is likely to encounter in real life. In contrast, most research on this topic has employed standardised, posed, extreme expressions (McClure, 2000), which are not necessarily representative of the kind of stimuli we normally have to work with. It did not necessarily follow that the developmental trends identified when using these different kinds of stimuli would be identical with what has been

reported in published literature; that said, the results of the current study are not radically different from those of past research (e.g. Rodger et al, 2015). This matter is discussed below.

Another key strength of the current research is that children were required to make judgments of facial expressions under a condition in which we can categorically determine whether they are correct or incorrect. Arguably, we do not know whether or not participants have made a correct judgment when we use standardised stimuli (i.e. 'Ekman faces') because: (a) the expression is merely a reflection of the target's notion of an angry expression, which is not necessarily the same as an expression they make when really angry; (b) the number of people who concur in judging that the expression is, say, 'angry' during a validation procedure is not necessarily a good indicator that the expression approximates to something that is natural. West and Kenny (2011) argue that one of the most elusive goals in research into person perception is to satisfy what they call the 'truth condition.' In other words, they argue that it is highly desirable to be able to classify social judgments as being categorially and objectively correct or incorrect, but to do so we need to know the true state (of the target) in relation to the participant's judgment.

The task presented here does indeed satisfy the truth condition and this is achieved by asking participants to guess which emoticon the target had chosen. The participant's choice can then be compared against an objective fact – we know for a fact which emoticon the target actually chose. The targets always chose an emoticon that correctly matched the expression they were shown and therefore the task involves social judgment that also satisfies the truth condition.

The task has some limitations as follows. First, it is difficult to explain any circumstance in which participants identify some target emotions more accurately than others. Because the target expressions are spontaneous mirrorings of expressions shown to them, we have no control over the strength of their mirroring and therefore no control over

the strength of the signal emitted by targets for interpretation by participants. For example, on finding that participants identified target expressions of happiness more accurately than other expressions, we do not know if this was because targets mirrored happiness more strongly than they mirrored other expressions or whether it is because irrespective of the strength of signal, participants are better adapted to recognizing a happy face more than other expressions.

For this reason, it is difficult to make comparisons between the results of the current study and those reported in the literature. Nevertheless, there are noteworthy instances where findings presented here are consistent with trends reported previously. Rodger et al (2015) observed that the expression of disgust was difficult for children to detect, leading to a sharp contrast between the performance of children and adults. Moreover, Widen and Russell (2008; Widen & Russell, 2015) found that disgust is one of the most difficult emotions to recognize or understand and in consequence development is tardy with respect to this particular emotion. In the current study the same trend emerges: Children were far less accurate in identifying disgust compared with adults. According to Widen and Russell, findings such as these are surprising in that they challenge the view that early understanding of disgust should be expected on account of its obvious survival value. That said, we cannot be sure in the current study whether children's poor identification of disgust was because they had difficulty processing a disgusted expression in targets or whether they had difficulty interpreting the emoticon that represented disgust which could in turn have presented a barrier to mapping this emoticon onto the relevant set of target expressions.

A second limitation is that our data are not informative about the developmental onset of recognizing subtle and spontaneous facial expressions. Children in all age groups performed well in judging which expressions targets were viewing, including those aged 7 years. Would younger children perform well, such as 5-year-olds or even 3-year-olds?

Rodger et al (2015) demonstrated systematic performance in their task down to age 5 years and this, coupled with the reported good performance of 7-year-olds in the current study, gives reason to believe that 5-year-olds and perhaps even 3-year-olds would be able to recognize spontaneous and subtle facial expressions. However, pilot work for the current study suggested that children younger than 7 years had difficulty working with the computer application developed for presenting the stimuli.

A third limitation relates to the form of judgment required of participants, which might effectively boost their chances of success, thus leading us to overestimate their abilities. Because the task involved a multiway forced choice, participants could have used a process of elimination – negatively ruling out alternatives that they recognized as being incorrect rather than positively identifying the correct option. Such a task probably supports judgments, perhaps allowing performance to be more accurate than if children had made a free response (Nelson & Russell, 2016).

The task used in the current study has the form of 'retrodictive mindreading.' *Retrodiction*, a concept articulated by Gallese and Goldman (1998), is the opposite of *prediction* and, according to Teoh et al (2017), refers to instances where we see a sample of behaviour and then work backwards to infer the mental state that underlies that behaviour (the proximal cause) and in turn the event (something that might have happened in the world) that triggered the mental state (the distal cause). In the current study, while we cannot be certain about the status of the proximal cause of the target's behaviour, we can at least be absolutely certain about the distal cause: We can be certain, for example, that the target was viewing a disgusted expression, and if the participant selected an emoticon for 'disgust', then it would appear that they had made a retrodictive inference.

Does this qualify as evidence that the children could read the target's minds? If children had actually followed the instruction to infer which emoticon the target had chosen,

then one might make a case for arguing that they were indeed mindreading as defined by

Gallese and Goldman (1998). However, despite the wording of the instruction, children might

have simply resorted to selecting an emoticon that matches the target's expression. If so, it is

nevertheless impressive that they could classify the target expressions with a degree of

accuracy and the conclusions of this article are confined to that. In short, while we offer

evidence that speaks to children's ability to classify subtle spontaneous expressions, we fall

short of being able to conclude that children could read the minds of the targets, defined as

inferring the cause of the target's expression (that the targets were in turn viewing an

expression of a certain kind).

References

Barth, J. M., & Batiani, A. (1997). A Longitudinal Study of Emotion Recognition and PreSchool Children's Social Behaviour. *Merrill-Palmer Quarterly, 43*(1), 107-128.

Balaban, M. T. (1995). Affective Influences on Startle in Five-Month-Old Infants: Reactions to Facial Expressions of Emotion. *Child Development, 66*(1), 28-36.

Baron-Cohen, S., Joliffe, T., Mortimore, C., & Robertson, M. (1997). Another Advanced Test of Theory of Mind: Evidence from Very High Functioning Adults with Autism or Asperger Syndrome. *The Journal of Child Psychology and Psychiatry, 38*(7), 813-822.

Calvo, M. G., & Lundqvist, D. (2008). Facial expressions of emotion (KDEF): Identification under different display-duration conditions. *Behavior Research Methods, 40*(1), 109-115.

Carroll, J. A., & Russell, J. A. (1997). Facial expressions in Hollywood's protrayal of emotion. *Journal of Personality and Social Psychology, 72*(1), 164-176.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology, 76*(6), 893-910.

Dimberg, U., Thunberg, M. & Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychological Science*, *11*, 86-89.

Ekman , P. (1992). Are there basic emotions? *Psychological Review, 99*(3), 550-553.

Farroni, T., Menon, E., Rigato, S. & Johnson, M.H. (2007). The perception of facial expressions in newborns. *European Journal of Developmental Psychology*, *4*, 2-13.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences, 2*(12), 493-501.

Hall, J. A. (1984). Nonverbal sex differences: Communication accuracy and expressive style. Baltimore: John Hopkins University Press.

Herba, C., & Phillips, M. (2004). Annotation: Development of facial expression recognition from childhood to adolescence: behavioural and neurological perspectives. *The Journal of Child Psychology and Psychiatry, 45*(7), 1185-1198.

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*(1), 18-23.

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

Lappänen, 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., & Ohman, A. (1998). The Karolinska Directed Emotional Faces – KDEF. In Department of Clinical Neuroscience, Psychology section, Karolinska Institutet.

McClure, E. B. (2000). A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychological Bulletin, 126*(3), 424-453.

Montagne, B., Kessels, R.P.C., Frigerio, E., de Haan, E.H. & Perrett, D.I. (2005). Sex differences in the perception of affective facial expressions: Do men really lack emotional sensitivity? *Cognitive Processing*, *6*, 136-141.

Nelson, C.A. & Russell, J.A. (2016). Building emotion categories: Children use a process of elimination when they encounter novel expressions. *Journal of Experimental Child Psychology*, *151*, 120-130.

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1-2), 8-13.

Pillai, D., Sheppard, E., & Mitchell, P. (2012). Can People Guess What Happened to Others

from Their Reactions? *PLOS One, 7*(11), e49859.

Pillai, D., Sheppard, E., Ropar, D., Marsh, L., Pearson, A., & Mitchell, P. (2014). Using
Other Minds as a Window Onto the World: Guessing What Happened from Clues in
Behaviour. *Journal of Autism and Developmental Disorders., 44*(10), 2430-2439.

Rodger, H., Vizioli, L., Ouyang, X., & Caldara, R. (2015). Mapping the development of
facial expression recognition. *Developmental Science, 18*(6), 929-939.

Sato, W. & Yoshikawa, S. (2007). Spontaneous facial mimicry in response to dynamic facial
expressions. *Cognition*, *104*, 1-18.

Stel, M., & Vonk, R. (2010). Mimicry in social interaction: Benefits for mimickers,
mimickees, and their interaction. *British Journal of Psychology, 101*(2), 311-323.

Teoh, Y., Wallis, E., Stephen, I.D. & Mitchell, P. (2017). Seeing the world through other
minds: Inferring social context from behaviour. *Cognition*, *159*, 48-60.

Theurel, A., Witt, A., Malsert, J., Lejeune, F., Fiorentini, C., Barisnikov, K. & Gentaz, E.
(2016). The integration of visual context information in facial emotion recognition in
5- to 15-year-olds. *Journal f Experimental Child Psychology*, *150*, 252-271.

Trantacosta, C.J. & Izard, C.E. (2007). Kindergarten children's emotion competence as a
predictor of their academic competence in first grade. *Emotion*, *7*, 77-88.

West, T. V., & Kenny, D. A. (2011). The truth and bias model of judgment. *Psychological
Review, 118*, 357 - 378.

Widen, S.C. & Russell, J.A. (2008). Children's and adults' understanding of the ''disgust
face." *Cognition and Emotion*, *22*, 1513-1541.

Widen, S.C. & Russell, J.A. (2008). Do dynamic facial expressions convey emotions to
children better than do static ones? *Journal of Cognition and Development*, *16*, 802-
811.

Table 1. Confusion matrix showing total sum of responses for each emoticon for each year group (7-, 9- and 11-year-olds). A= happy, B= Sad, C= Angry, D = Shocked and E = Disgusted.

		Emoticon chosen by child participant														
		<u>A</u>			<u>B</u>			<u>C</u>			<u>D</u>			<u>E</u>		
		7	9	11	7	9	11	7	9	11	7	9	11	7	9	11
Emoticon Chosen by target	A	174	203	205	60	37	38	67	78	49	30	45	19	24	25	20
	B	52	59	66	107	121	107	105	132	90	60	40	27	29	37	39
	C	47	44	36	58	48	44	139	170	142	77	80	60	34	46	47
	D	26	39	21	64	54	53	143	148	109	95	106	94	25	35	48
	E	52	54	50	63	42	43	98	124	75	94	108	82	57	60	81

Shaded areas indicate where highest values are expected due to these reflecting the correct response.

Figures

Figure 1. Five emoticons (happy, sad, angry, shocked and disgusted).

Figure 2. Screen shots of some of the target videos presented to the participants. Top: Left to right targets viewing expressions that were Happy, Sad and Angry. Bottom: Left to right, targets viewing expressions that were Shocked and Disgusted.
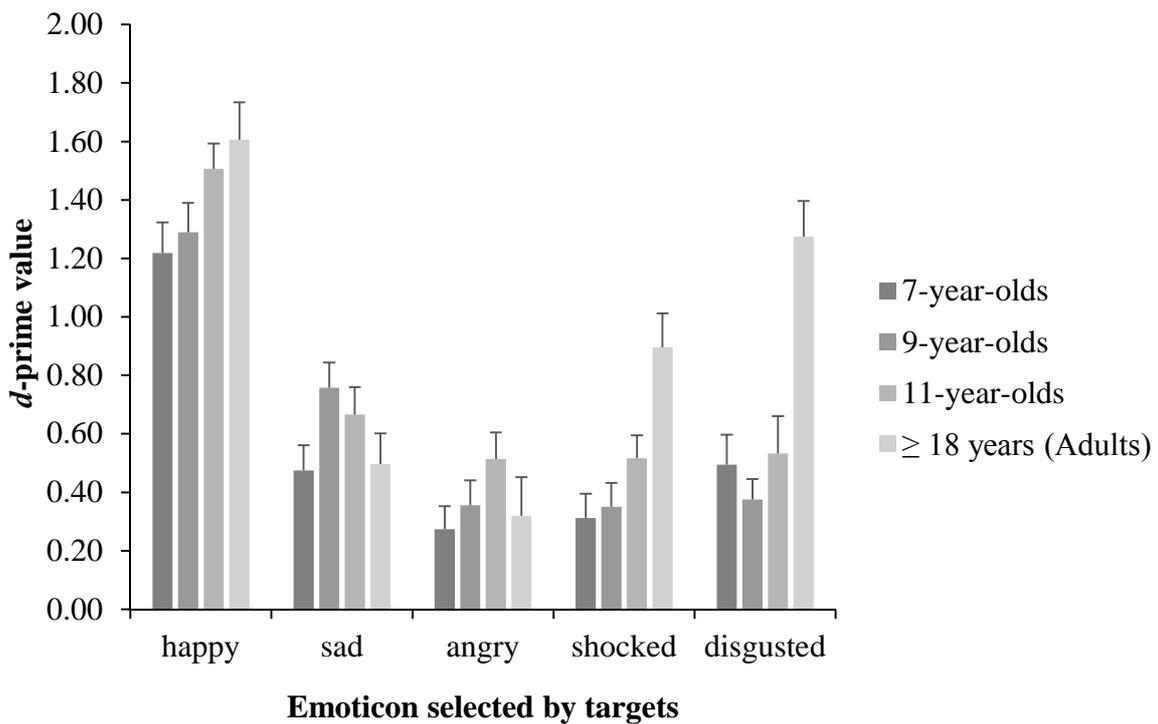
Figure 3. Means (and Standard Error) of *d*-prime values that represent children's (7-year-olds, 9-year-olds, 11-year-olds) and adult's accuracy in determining which emoticon (happy, sad, angry, shocked and disgusted) had been chosen by targets. All 20 means were significantly greater than 0 ($p \leq .01$ in all cases) according to one sample *t*-tests.