Journal of Experimental Social Psychology 60 (2015) 150–156

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

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

Face-trait inferences show robust child–adult agreement: Evidence from three types of faces

E.J. Cogsdill *, M.R. Banaji

Harvard University

HIGHLIGHTS

- Face-to-trait inferences are robust early in childhood and generalize to diverse facial stimuli.
- Child–adult agreement is observed for three different types of face stimuli: human adults, children, and rhesus macaques.
- This agreement replicates when a larger number of face images is used.

ARTICLE INFO

Article history:
Received 13 November 2014
Revised 28 April 2015
Accepted 1 May 2015
Available online 27 May 2015

Keywords:
Social cognition
Social cognitive development
Person judgment
First impressions
Trait impressions
Face processing

ABSTRACT

Humans rapidly and automatically use facial appearance to attribute personality traits (“trustworthy,” “competent”). To what extent is this face-to-trait attribution learned gradually across development versus early in childhood? Here, we demonstrate that child–adult concordance occurs even when faces should minimize agreement: natural (not computer-generated) adult faces; less developed children's faces; and perceptually unfamiliar monkey faces. In Study 1, 3- to 12-year-olds and adults selected “nice/mean” faces among pairs with a priori “nice-mean” ratings. Significant cross-age consensus emerged for all three face types. Study 2 replicated this result using an improved procedure in which 44–48 faces appeared in randomized pairs. This converging evidence supports the idea that complex forms of social cognition – allowing perceivers to believe they can derive personality from faces – emerge early in childhood, a finding that calls for new procedures to detect this central facet of cognition earlier in life.

1. Introduction

Faces communicate a wealth of information about other people, and specialized mechanisms have evolved for the purpose of using faces to make inferences about others (Todorov, Said, Engell, & Oosterhof, 2008; Zebrowitz, 1997). The past decade of research in social cognition has clearly demonstrated that not only do we decode basic group membership of individuals upon viewing their faces, but we also attribute personality traits to them based on their physiognomy. Whether accurate or not, judgments of character traits like trustworthiness, competence, and dominance are routinely and consensually offered. Prior work has shown that judgments about others' personality from facial appearance form rapidly (Rule & Ambady, 2008a) and change little over time (Willis & Todorov, 2006). Importantly, face-based trait inferences are consistent both within and across cultures, and predict important outcomes such as election results (Ballew & Todorov, 2007; Todorov, Mandisodza, Goren, & Hall, 2005) and career success (Rule & Ambady, 2008b, 2011). Face-to-trait judgments have thus emerged as a major source of interest for experimental psychologists seeking to understand the social cognitive bases of social perception.

We know a great deal about how adults attribute personality to others based on their physical appearance. For example, perceivers use static facial characteristics to inform decisions about personality traits and even complex beliefs such as religious affiliations (Rule, Garrett, & Ambady, 2010). However, little is known about the development of the fundamental tendency to infer traits such as “trustworthy” or “competent” from facial appearance alone. Judgments of interpersonal warmth are uniquely important given the wealth of evidence from evolutionary social cognition research suggesting that such judgments are fundamental among dimensions of interpersonal perception (Fiske, Cuddy, & Glick, 2007). Furthermore, the primacy of warmth evaluation speaks to the importance of selective pressures in guiding human tendencies in forming impressions about one another (Schaller, 2008). Understanding the developmental process of such judgments can therefore inform whether face-to-trait inferences, and inferences of warmth in particular (i.e., “nice” or “mean”), represent a fundamental and deep-rooted

* Corresponding author.
E-mail addresses: emily.cogsdill@gmail.com (E.J. Cogsdill), mahzarin_banaji@harvard.edu (M.R. Banaji).

http://dx.doi.org/10.1016/j.jesp.2015.05.007
0022-1031/

element of social cognition, in which case they should be observed relatively early in development, or whether they instead require protracted social learning and experience to be cultivated.

Previous work has shown that even infants possess the roots of face perception (Cassia, Turati, & Simion, 2004; Pascalis, de Haan, & Nelson, 2002) and even demonstrate certain preferences based on facial attractiveness (e.g., Langlois et al., 2000) and group membership in domains like gender (Quinn, Yahr, Kuhn, Slater, & Pascalis, 2002) and ethnicity (Bar-Haim, Ziv, Lamy, & Hodes, 2006; Kelly et al., 2005). Recent work has also begun to suggest that face-to-trait inferences may emerge early in development. When asked to choose faces possessing certain traits, such as “nice,” “strong,” or “smart,” children as young as 3–4 years of age tended to choose the same types of faces as adults (Cogsdill, Todorov, Spelke, & Banaji, 2014). This cross-age consensus was especially pronounced when children were asked to make basic evaluative judgments (i.e., “Which of these people is very nice/mean?”), and all participants did so readily for all of the types of faces. Thus, the face perception skills and early social preferences that have been observed in infants also appear to manifest in face-based trait inferences that emerge by the preschool years, if not even earlier.

However, the research to date is limited in a variety of ways. Cogsdill et al. (2014) used only computer-generated face stimuli that were designed to strongly elicit judgments of high or low trustworthiness, dominance, and competence (Oosterhof & Todorov, 2008). Besides being computer-generated rather than natural in appearance, these faces were also specifically selected for their extreme representation of each trait. Natural faces representing a milder range of features may fail to elicit the same attributions from children.

To test the strength of face-to-trait inference across development, the present work measures cross-age consensus in large numbers of children and adults when evaluating naturalistic human faces. If children fail to produce adult-like responses while viewing natural human faces, this would suggest that Cogsdill et al. (2014) overstated the degree to which children form adult-like trait attributions in response to faces. If, on the other hand, the developmental continuity that was observed with more extreme computer-generated faces is mirrored here, this would provide compelling supporting evidence that child–adult agreement in face-to-trait inference can emerge in the natural world.

In addition to using naturalistic faces of adult targets as stimuli, we also ask participants to judge faces belonging to children, whose underdeveloped features are less likely to suggest strong personality traits. By using faces belonging to children, we can therefore begin to explore whether the features that perceivers interpret as signaling “niceness” or “meanness” even when in weak form allow social detection early in life or whether such weak facial signals allow detection only later in development, particularly after the onset of puberty.

Yet another conservative test of the development of face-trait inferences can be achieved by using faces that are novel to both children and adults. Faces belonging to non-human primates provide such a set. Because participants are unfamiliar with monkey faces, these stimuli will reveal whether perceptual experience affects the development of consensual face-to-trait judgments. If we observe agreement between the judgments that children and adults apply to such unfamiliar stimuli, this would further corroborate the robustness of face-to-trait inferences even in early childhood.

Monkey faces confer an additional advantage in that they are less susceptible to the culture-specific trait inferences that may guide judgments of human faces. Even though the adult human faces we used were deliberately altered to appear relatively homogenous in appearance (through manipulations described below), some stereotypes may still influence judgments of people from different cultures. For example, even though all faces belong to European targets, some may to belong to different groups within that ethnic designation, a fact that some perceivers might be able to recognize. The lack of culture-specific biases associated with monkey faces thus affords the ability to test the development of face-to-trait inference devoid of such influences.

An additional limitation of prior work lies in the limited sample sizes and truncated age ranges that are typically used. Most studies of social cognitive development do not include sufficiently large samples of children to be able to conduct meaningful correlational analyses to track changes in social judgments across childhood. We set a priori standards for large samples to allow such analyses, with predetermined samples of $n = 100$ children and adults viewing each of the three types of faces (adults, children, and monkeys), for combined samples of $N = 600$ in both experiments.

2. Experiment 1

2.1. Method

2.1.1. Participants

A total of 600 individuals participated in Experiment 1. Half of this sample consisted of children between the ages of 3 and 13 who participated at the Boston Children's Museum ($M = 6$ years 11 months, $SD = 2$ years 3 months), and the other half consisted of ($n = 300$) adults ages 18 to 61 who participated online through SocialSci.com (M age = 27.93 years, $SD = 9.56$). Both the child and adult samples were predominantly White, with 66% of children and 71% of adults identifying as White. Roughly equal numbers of males and females participated among both children (56% female; 1 did not specify) and adults (56% female; 9 unspecified).

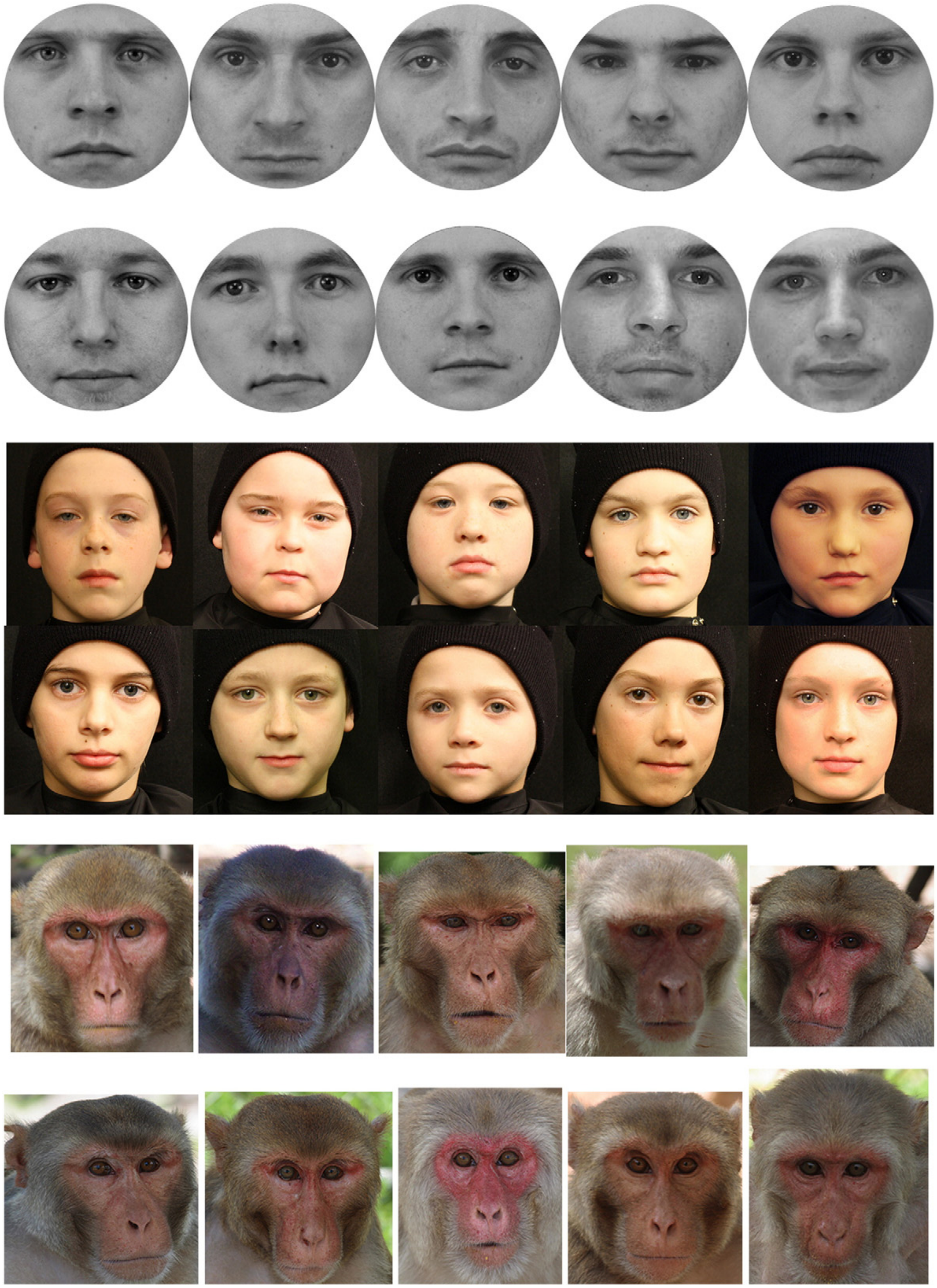
2.1.2. Stimuli

Experiment 1 used three different types of face stimuli: images of adult human faces, children's faces, and rhesus macaque monkey faces. For each set, a separate pilot study was conducted with adults participating online (www.socialsci.com). Three independent samples of 100 adults rated the three sets of faces in terms of how “Nice” they appeared, on a scale from 1 (extremely mean) to 5 (extremely nice). Pilot ratings were obtained in this manner for a total of 45 adult faces, 49 child faces, and 20 monkey faces.

The three types of faces differed significantly in their visual characteristics. Adult faces were obtained from a set of photos taken of emotionally neutral, White male actors in Karolinska, Sweden (Lundqvist, Flykt, & Ohman, 1998; Fig. 1). All male photos in this set were cropped into a circular image such that participants were forced to rely only on internal facial features when rendering their judgments. Photos were converted to grayscale and image brightness was standardized across the set. These steps were taken to deliberately render the set of adult faces as homogeneous as possible, thereby further increasing the difficulty of the task in order to provide a more conservative test of correspondence between children's and adults' judgments.

Children's faces were obtained from the Dartmouth Database of Children's Faces (Dalrymple, Gomez, & Duchaine, 2013; Fig. 2). While the full database includes faces of both males and females between the ages of 5 and 16, only a subset of 49 male faces was piloted for use in this study. The ages of children included in the final set of stimuli for Experiment 1 was more narrow than that of the full database, ranging from only 8 to 12 years of age ($M = 9.60$, $SD = 1.58$). As with adult faces, only male faces were included so as to avoid effects driven by gender stereotypes and to maintain consistency across the face conditions. Children were photographed wearing black shirts and caps such that their ears, hair, and clothing were not visible. Unlike the adult faces, however, some external features of children's faces remained visible, and their faces were presented in full color with variable brightness.

Rhesus macaque faces were presented in the most straightforward manner, with monkeys' entire heads simply being cropped out of larger images provided by Laurie Santos' Comparative Cognition Laboratory (Fig. 3). A set of 20 faces was presented in full color in Experiment 1 and in grayscale for Experiment 2; the presence of color did not appear to significantly alter participants' responses.



Figs. 1–3. Examples of stimuli depicting adult, children, and monkey faces, with "mean" faces (based on pilot ratings) in the top row and "nice" faces on the bottom row.

Using the pilot ratings collected for each stimulus type, five "nice" and five "mean" faces were selected for use in each of the three face conditions (Adult, Child, and Monkey Faces). Faces were selected from across the full range of pilot ratings such that those ratings were matched across all three stimulus sets; i.e., for every face in a given set, we selected a face in both of the other sets whose average pilot rating closely matched that face. This feature of the stimuli allows us to directly compare results obtained using the three face sets.

2.1.3. Procedure

Children completed the experiment by pointing to a laptop while being seated facing the pairs of stimulus faces. They first completed a practice round consisting of three trials in which the instruction was simply to point to the larger or smaller face, a task that all children completed easily, ensuring that children understood the task and were pointing clearly to the correct face on the screen. No practice trials were included for the adult version, which participants completed on www.socialsci.com on their own computers.

For the main experiment, participants viewed pairs of faces and answered the question, "which of these [people/children/monkeys] is very [nice/mean]?" Each face pair consisted of one "nice" and one "mean" face based on pilot ratings, with all combinations of five nice and five mean faces yielding a total of 25 trials. For all participants, the "nice" face appeared on the left side of the screen in either 12 or 13 of 25 trials. The order of face pairs was randomized such that no face appeared in consecutive trials, and prompts were randomized so that the anticipated response was on the left or right sides with approximately equal frequency (12 or 13 trials). All participants saw faces from only one stimulus set, making Face Type a between-subjects variable (Adult vs. Children vs. Monkey faces). Children were assured that there were no right or wrong answers.

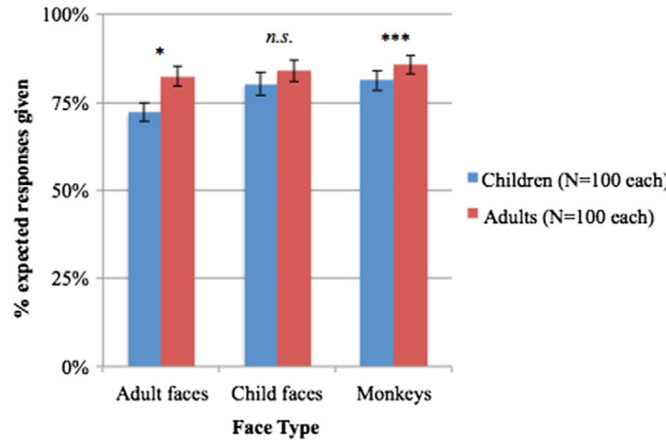


Fig. 4. Overall patterns of response from Experiment 1. Error bars represent 95% confidence intervals.

2.2. Results[1]

Overall patterns of response were quite consistent (Fig. 4). Participants provided anticipated responses in 81% of all trials, a rate that was significantly higher than chance (50%; $t(599) = 49.19$, $p < .001$, $d = 4.02$). This consistency was similarly high when data were separated by Age Group, with children and adults providing expected responses in 78% and 84% of all trials, respectively. Both children and adults were most consistent when judging monkey faces, for which they provided expected responses in 81% and 86% of trials respectively. Although adults were similarly consistent when judging faces belonging to children (84%) and other adults (82%), child participants were less reliable when judging those faces, giving anticipated responses in 80% of trials when viewing children's faces and 72% of trials when viewing adult faces. All judgments by all age groups greatly exceeded chance responding at 50% (all $ts(99) > 16.04$, all $ps < .001$, all $ds > 3.22$).

Data were first analyzed in the aggregate using a 3 × 2 ANOVA, with Face Condition as the first factor (Adult, Child, and Monkey faces) and Age Group as the second factor (Child vs. Adult participants). This analysis revealed a significant effect of Age Group, with adults providing more expected responses than children across the full set of data, $F(1,594) = 25.74$, $p < .001$, $\eta^2 = .042$. The effect of Face Condition was also significant, $F(2,594) = 9.46$, $p < .001$, $\eta^2 = .031$. The Age Group × Face Condition interaction was marginally significant, $F(2,594) = 2.80$, $p = .062$, $\eta^2 = .009$. Thus, adults and children differed in their responses, and the type of face being judged affected patterns of response, with the interaction between the two approaching significance.

The significant main effects of Face Condition and Age Group were further analyzed. First, Sidak-corrected post-hoc analyses on the original ANOVA compared response patterns between the three Face Conditions, revealing that responses to adult faces (77%) were less reliable than those generated in response to faces belonging to children (82%, $p < .01$, $d = 0.30$) and monkeys (84%, $p < .001$, $d = 0.42$), which were not significantly different from one another ($p = 0.68$, $d = 0.10$). Next, independent samples t-tests compared responses between adults and children in each of the three Face Conditions (Adult, Child, and Monkey Faces). Here, the largest difference in responses between adults and children was observed for Adult Faces, in which adults provided the expected responses in 82% of all trials, whereas children did so in only 72% of trials, $t(198) = 5.07$, $p < .001$, $d = 0.72$. In the Monkey Faces condition, adults provided the expected responses in 86% of all trials, while children did so in 81% of trials ($t(198) = 2.26$, $p < .05$, $d = 0.32$). Child–adult agreement was greatest in the Child Faces condition, in which adults provided expected responses in 83.9% and children in 80.1% of all trials, a difference that did not reach statistical significance ($t(198) = 1.64$, *n.s.*). The marginally significant Age Group × Face Condition interaction therefore resulted from the fact that cross-age agreement differed depending on the face being perceived, with the greatest consensus emerging in response to faces belonging to children.

The relatively large samples of children we collected enabled us to conduct correlational analyses to explore the developmental trajectories of judgments in the three Face Conditions. To do this, we measured Pearson correlations to study the effects of age on patterns of response among child participants, with adult participants' responses being excluded from analysis. Here, the size of each correlation (i.e., the effect size) represents the magnitude of the effect of development on consensus of response. When data were collapsed across Face Condition ($N = 300$), the relationship between age and consensus was highly significant, $r = 0.30$, $p < .001$. All three Face Conditions yielded nearly identical effect sizes when analyzed separately (all $Ns = 100$, $rs = 0.33$, $ps < .001$), and further analyses using Fisher r-to-z transformations confirmed that the differences between these correlations were not significant (all $zs < 0.02$, all $ps > 0.98$).

2.2.1. Discussion

The foregoing analyses suggest two key conclusions. First, the overall percentages of expected responses given indicate that all participants showed significant consensus, with children's responses predominantly matching those of adults. This first finding converges with that of Cogsdill et al. (2014), and critically demonstrates that such high levels of agreement can be achieved not only with computer-generated faces that were designed to elicit certain trait judgments, but also with a variety of naturalistic stimuli as well. The fact that such high levels of consensus were observed for three new types of stimuli, which varied both in terms of the type of subject being portrayed (adults, children, and monkeys) as well as in a number of visual characteristics (e.g., color vs. grayscale; presence vs. absence of external facial features), suggests that this cross-age consensus is based on judgments that are sufficiently robust early in childhood to generalize to a variety of face stimuli. The second key finding is that, although significant consensus was observed for both children and adults in all three Face Type conditions, this consensus does increase with age, suggesting that the roots of these face-based evaluations are both clearly present early in childhood and continue to develop nuances throughout the age range tested.

Surprisingly, however, correlational analyses measuring the effects of development on consensus among children yielded practically identical effect sizes for the three types of stimuli used. This striking convergence further suggests the possibility that the significant interaction observed between Age Group × Face Condition may be due to developmental changes occurring during adolescence, during the ages between

[1] All data are publicly available through the Open Science Framework at the following URL: https://osf.io/5x9iq/.

those of the oldest ages of children and the youngest adults we tested. The possibility that face-based evaluations might actually "mature" over a prolonged period of time over the course of development, even extending into adolescence, is consistent with recent work on face processing suggesting that face recognition (e.g., Lawrence et al., 2008) and that the neural regions implicated in face processing also develop in a highly prolonged fashion, reaching maturity in the adolescent years or later (e.g., Scherf, Behrmann, Humphreys, & Luna, 2007). Such work suggests the possibility that the ability to consistently attribute traits to faces with subtle variations in appearance may follow a highly prolonged period of development as the full complement of face processing abilities matures.

This consistency of developmental effects across the three Face Type conditions clearly demonstrates that the type of face being judged was relatively unimportant in guiding the consistency of children's responses. In other words, it does not seem to matter exactly what kind of face a participant is observing. Whatever intuitions children use to guide their face judgments appear to be present early in life, develop across childhood (and perhaps even into adolescence), and are readily applied to diverse types of face stimuli, owing to their robust nature early in life.

3. Experiment 2

Although Experiment 1 showed unambiguously that consensus is both highly significant for all three types of stimuli tested and continues to develop across childhood (and potentially through adolescence and adulthood as well), the experimental design was limited by the relatively small number of face stimuli that were used. Specifically, each Face Type consisted of only 10 face identities, with five "nice" and five "mean" faces being selected for each condition based on pilot ratings. Experiment 2 sought to replicate and extend the findings of Experiment 1 using an improved procedure in which all participants viewed the entire sets of 44–48 faces, which were paired randomly in each trial and for each participant. The inclusion of a larger number of faces in each set produced a more robust design in which overall patterns of results would be less susceptible to variability caused by any individual face.

3.1. Method

3.1.1. Stimuli

Instead of using pilot ratings to select sets of five "nice" and five "mean" faces from each set of faces, as was done in Experiment 1, Experiment 2 used all faces belonging to each set, with two exceptions. In order to create sets containing even numbers of faces, one adult and one child face was removed. The resulting set of 48 male children's faces ranged in age from 5 to 14 years ($M = 9.25$, $SD = 1.86$). We converted the monkey faces to grayscale to minimize the effects of color on children's judgments and added 26 new faces to increase the size of the set to 46 faces so that it would more closely match those of the two sets of human faces. We did not collect pilot ratings for these additional monkey faces. Experiment 2 thus used a total of 44 adult faces, 48 child faces, and 46 monkey faces.

3.1.2. Procedure

Adult participants completed Experiment 2 on web sites that presented stimuli and recorded responses from their own computers. These web-based experiments enabled us to use a more robust procedure in which all faces were paired randomly and appeared only once throughout the experiment, which allowed for a larger number of faces to be included. Adults accessed the experiments online through Amazon Mechanical Turk. As in Experiment 1, children viewed the experiments on a computer while seated next to the experimenter.

The experimental procedure was identical for all three Face Types, with the sole exception consisting of the number of trials presented, with 22, 24, and 23 trials of Adult, Child and Monkey faces, respectively. Children's versions of the experiment began with the same three practice trials that were used in Experiment 1. During the main portion of the experiment, children were instructed to simply point to the "nice" face in each trial. The experimenter further explained that children could simply keep pointing to the nicer face in each pair when new faces would appear on the screen, rather than waiting for the experimenter to ask them which one was nicer. This was done to eliminate the need to repeat the same question in all 22–24 trials.

During each trial, a pair of faces would appear on the screen. In the first trial, children were prompted to "Start pointing to the nicer faces," after which they would point to one of the faces on the screen. The experimenter would click on the face they selected, upon which the experiment would automatically advance to the next trial by displaying a new pair of faces. This continued for all 22–24 trials until all faces had appeared, at which time the child would be thanked and offered their choice of sticker.

Children's face selections were recorded automatically in the accompanying database, and the experimenter separately recorded their date of birth, sex, and race, with the latter two fields being optional demographic fields on the parental consent form. The adult version of the experiment included a demographic form at the end of the experiment that similarly prompted participants to provide their age in years (required), as well as their sex and racial identification (optional).

3.1.3. Results

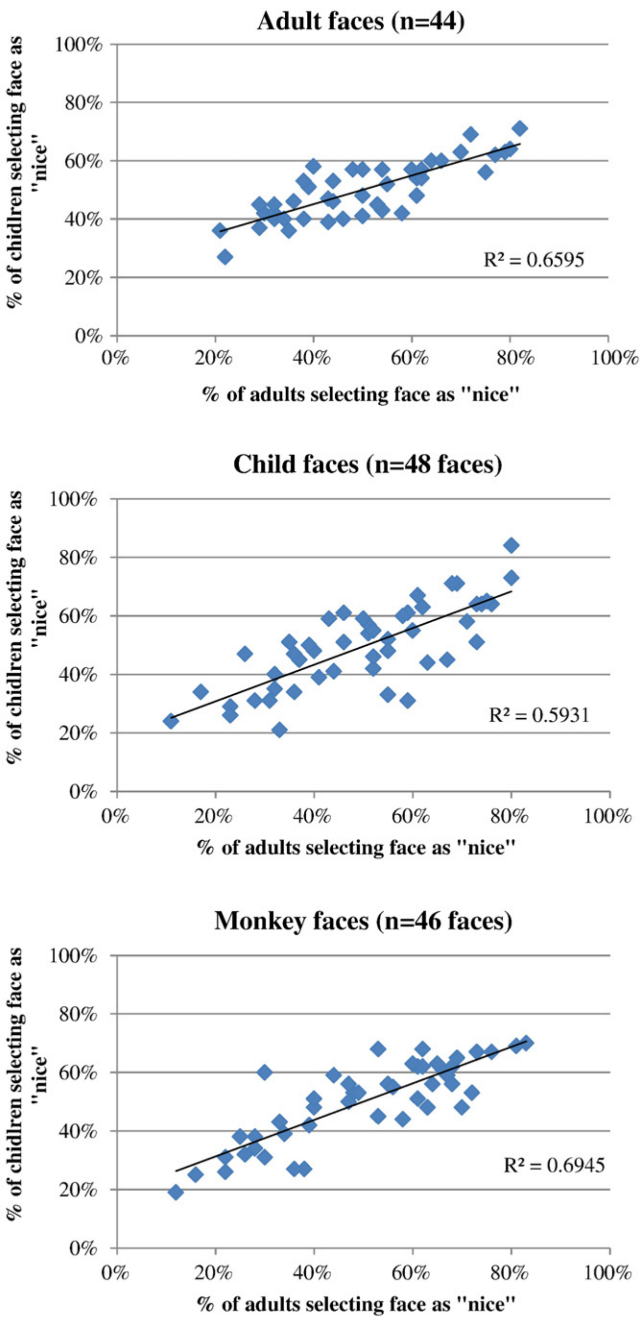
For each face in each stimulus set, the percentage of children and adults who selected the face as the "nice" one was recorded. Consensus was thus measured as the strength of the relationship between children's and adult's frequencies of categorizing individual faces as "nice." These associations were measured using Pearson correlations between these frequencies for all faces in the three sets, which represent the degree of correspondence between children's and adult's selections of faces belonging to adults ($N = 44$ faces, Fig. 5), children ($N = 48$ faces, Fig. 6), and monkeys ($N = 46$ images, Fig. 7). These frequencies were obtained from samples of $N = 100$ adults and $N = 100$ children for each of the three face types (overall $N = 600$).

This analysis reveals strong consensus between children's and adult's face categorizations for all three types of stimuli, all rs > .77. Fisher r-to-z transformations revealed no differences in the strengths of correlations between the three face types (all zs < 1.16, all ps > 0.25). Children ages 3–13 thus showed comparable levels of consensus with adults in terms of their categorization of faces belonging to adults, children, and even monkeys.

Further analyses were conducted to observe how this child–adult consensus might develop with age among child participants. To do this, we systematically excluded children at each age (starting by excluding 13-year-olds, then 12-year-olds, and so on) and measured the correlations that result from comparing adult categorization frequencies to these progressively smaller and younger groups of children. This analysis allows us to investigate the lower limits of children's tendency to generate adult-like face categorizations.

The results of this analysis, which are summarized in Tables 1–3, show that a significant correlation between children's and adult's face categorizations persists even when a large portion of the child samples are excluded from analysis. Specifically, for categorization of adult and child faces, correlations remained highly significant even when the samples were reduced to include only 3–5 year olds, with only $n = 35$ children viewing adult faces ($r = 0.35$, $p < .01$; Table 1) and $n = 28$ viewing children's faces ($r = 0.61$, $p < .001$; Table 2). When viewing monkey faces, this relationship remained significant even when only the responses of 3–4 year olds were compared to those of adults ($n = 18$ children, $r = 0.45$, $p < .001$; Table 3).

This analysis shows that the consensus we observe between children's and adult's face categorizations emerges very early in development, with children as young as 3–5 years of age showing significant consensus when judging human faces and even faces of nonhuman



Figs. 5–7. Results from Experiment 2. Plots display the percentages of children and adults who selected each of the children, adult, and monkey faces as being "nice."

primates. Additionally, since the set of children's faces in Experiment 2 represented a much broader range of ages (5–14 years old) than those in Experiment 1 (8–12 years old), this provides an even clearer demonstration of the broad generalizability of face-to-trait inferences.

4. General discussion

The present research suggests that using faces to infer personality characteristics, irrespective of accuracy, begins to take root early in childhood and is strongly generalizable even when the faces provide weak signals. As such, face-to-trait inferences must be considered an even more fundamental aspect of social cognition than was previously known, particularly when they concern inferences of interpersonal warmth, which is known to be a fundamentally important dimension of social perception. Critically, the experiments described above show that cross-age consensus exists even when faces make it difficult to signal personality, and that these judgments are overall quite consistent and robust across different face types. An intriguing complication of this basic finding is that not only are such judgments robust early in childhood, but they also continue to evolve over a prolonged span of development at least into adolescence. Thus, while face-to-trait inference appears to be present in surprisingly adult-like form early in childhood, it equally surprisingly also appears to continue to develop across childhood, as is the case for other aspects of brain development as well.

Tables 1–3
Additional analyses from Experiment 2. Tables display correlations between children's and adults' categorizations of adult, children, and monkey faces (respectively) with progressively younger samples of children.

Children's maximum age	*N* children	Child/adult correlation coefficients (*N* = 44 faces)	t-Value	*p*-Value (df = 46, two-tailed)
12	100	0.81	9.02	<0.01
11	98	0.81	9.08	<0.01
10	94	0.81	9.04	<0.01
9	87	0.80	8.58	<0.01
8	81	0.76	7.70	<0.01
7	74	0.76	7.70	<0.01
6	53	0.62	5.12	<0.01
5	35	0.44	3.13	<0.01
4	19	0.18	1.20	0.24

Children's maximum age	*N* children	Child/adult correlation coefficients (*N* = 48 faces)	t-value	p-Value (df = 46, two-tailed)
13	100	0.78	8.40	<0.01
12	97	0.78	8.36	<0.01
11	93	0.78	8.40	<0.01
10	92	0.78	8.49	<0.01
9	82	0.77	8.20	<0.01
8	68	0.78	8.46	<0.01
7	54	0.75	7.76	<0.01
6	44	0.68	6.32	<0.01
5	28	0.61	5.22	<0.01
4	13	0.20	1.36	0.18

Children's maximum age	*N* children	Child/adult correlation coefficients (*N* = 46 faces)	t-Value	p-Value (df = 44, two-tailed)
12	100	0.83	10.00	<0.01
11	99	0.84	10.31	<0.01
10	93	0.84	10.34	<0.01
9	87	0.83	9.81	<0.01
8	73	0.82	9.43	<0.01
7	62	0.78	8.18	<0.01
6	48	0.72	6.92	<0.01
5	31	0.59	4.82	<0.01
4	18	0.45	3.33	<0.01
3	7	−0.05	−0.32	0.75

The use of naturalistic stimuli makes the present findings a particularly meaningful contribution. In particular, by using stimuli with more natural and subtle variation in facial features than was present in the computer-generated faces used in Cogsdill et al. (2014), we rendered the task more difficult. The inclusion of children's faces and monkey faces further contributed to this difficulty by providing targets with features that were not fully developed (as with children's faces) as well as targets with whom participants lacked extensive perceptual experience (i.e., monkey faces). If children's face-based evaluations were only tenuous, we should have expected them to manifest to a lesser extent, or potentially not at all, in response to these new stimuli. The convergent findings that emerged across all three types of faces lead us to conclude instead that children's tendency to evaluate faces in adult-like patterns is in fact a highly robust one that is readily generalized to many different types of faces.

Future work will be necessary to further understand both the development of the facial features that signal personality traits and the role of perceptual experience in guiding the development of consensus in making these judgments. By identifying the earliest age at which people are subject to face-trait judgments both from other children and adults, we can better understand not only the roots of the judgments themselves but the earliest point at which judgments may begin to affect behaviors towards even young children or infants. Facial features that guide face-to-trait judgments may emerge very early in development, and this could lead to social consequences with cascading effects across the lifespan.

In addition, the finding in Experiment 1 that the difference in consistency between adults and children was greater when attributing "nice/mean" to adult human faces as opposed to monkey faces suggests the possibility that developmental changes in perceiving traits in human faces might be driven by a gradual accumulation of perceptual experience with those faces. Future research should seek to clarify this result by studying older adolescent participants (ages 13–17), to assess whether developmental changes do indeed persist throughout adolescence, as well as to directly measure the effects of perceptual experience by manipulating participants' exposure to unfamiliar face stimuli. If perceptual experience does influence the frequency with which children dispatch adult-like judgments to faces, this would suggest a possible mechanism by which judgments continue to become increasingly consistent throughout adolescence and into adulthood.

The results of Experiments 1 and 2 underscore the fundamental nature of face-based trait impressions in social interactions at all ages. This propensity is therefore one that both emerges in a highly robust and generalizable fashion very early in life, yet also continues to develop into a fully adult-like state across the lifespan. Face-to-trait inferences were robust at the earliest ages tested, a fact that highlights the need for researchers to develop methods for studying even younger participants, even as early as in infancy. Such work will be necessary to fully understand the earliest origins of this central aspect of human social cognition.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jesp.2015.05.007.

References

Ballew, A., & Todorov, A. (2007). Predicting political elections from rapid and unreflective face judgments. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(46), 17948-43.

Bar-Haim, Y., Ziv, T., Lamy, D., & Hodes, R.M. (2006). Nature and nurture in own-race face processing. *Psychological Science*, *17*(2), 159–163.

Cassia, V.M., Turati, C., & Simion, F. (2004). Can a nonspecific bias toward top-heavy patterns explain newborns' face preference? *Psychological Science*, *15*, 379.

Cogsdill, E.J., Todorov, A., Spelke, E.S., & Banaji, M.R. (2014). Inferring character from faces: A developmental study. *Psychological Science*, *25*(5), 1132–1139.

Dalrymple, K.A., Gomez, J., & Duchaine, B. (2013). The Dartmouth database of children's faces: Acquisition and validation of a new face stimulus set. *PLoS ONE*, *8*(11), e79131. http://dx.doi.org/10.1371/journal.pone.0079131.

Fiske, S.T., Cuddy, A.J.C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, *11*(2), 77–83. http://dx.doi.org/10.1016/j.tics.2006.11.005.

Kelly, D.J., Quinn, P.C., Slater, A.M., Lee, K., Gibson, A., Smith, M., et al. (2005). Three-month-olds, but not newborns, prefer own-race faces. *Developmental Science*, *8*, 31–36.

Langlois, J.H., Kalakanis, L., Rubenstein, A.J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin*, *126*, 390–423.

Lawrence, K.K., Bernstein, D.D., Pearson, R.R., Mandy, W.W., Campbell, R.R., & Skuse, D.D. (2008). Changing abilities in recognition of unfamiliar face photographs through childhood and adolescence: performance on a test of non-verbal immediate memory (Warrington RMF) from 6 to 16 years. *Journal of Neuropsychology*, *2*(1), 27–45. http://dx.doi.org/10.1348/174866407X231074.

Lundqvist, D., Flykt, A., & Ohman, A. (1998). *The Karolinska directed emotional faces [database of standardized facial images]*. S-171 76 Stockholm, Sweden: Psychology Section, Department of Clinical Neuroscience, Karolinska Hospital.

Oosterhof, N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(32), 11087–11092.

Pascalis, O., de Haan, M., & Nelson, C. a (2002). Is face processing species-specific during the first year of life? *Science*, *296*(5571), 1321–1323. http://dx.doi.org/10.1126/science.1070223.

Quinn, P., Yahr, J., Kuhn, A., Slater, A., & Pascalis, O. (2002). Representation of the gender of human faces by infants: A preference for female. *Perception*, *31*, 1109–1121.

Rule, N.O., & Ambady, N. (2008a). Brief exposures: Male sexual orientation is accurately perceived at 50 ms. *Journal of Experimental Social Psychology*, *44*, 1100–1105.

Rule, N.O., & Ambady, N. (2008b). The face of success: Inferences from chief executive officers' appearance predict company profits. *Psychological Science*, *19*, 109–111.

Rule, N.O., & Ambady, N. (2011). Judgments of power from college yearbook photos and later career success. *Social Psychological and Personality Science*, *2*(2), 154–158.

Rule, N.O., Garrett, J.V., & Ambady, N. (2010). On the perception of religious group membership from faces. *PLoS ONE*, *5*(12), e14241. http://dx.doi.org/10.1371/journal.pone.0014241.

Schaller, M. (2008). Evolutionary bases of first impressions. In N. Ambady, & J.J. Skowronski (Eds.), *First impressions* (pp. 15–34). New York, NY, US: Guilford Publications.

Scherf, K.S., Behrmann, M., Humphreys, K., & Luna, B. (2007). Visual category-selectivity for faces, places and objects emerges along different developmental trajectories. *Developmental Science*, *10*(4), F15–F30. http://dx.doi.org/10.1111/j.1467-7687.2007.00595.x.

Todorov, A., Mandisodza, A.N., Goren, A., & Hall, C.C. (2005). Inferences of competence from faces predict election outcomes. *Science*, *308*(5728), 1623–1626.

Todorov, A., Said, C.P., Engell, A.D., & Oosterhof, N.N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, *12*(12), 455–460. http://dx.doi.org/10.1016/j.tics.2008.10.001.

Willis, J., & Todorov, A. (2006). Making up your mind after a 100-ms exposure to a face. *Psychological Science*, *17*(7), 592–598.

Zebrowitz, L.A. (1997). *Reading faces: Window to the soul?* Boulder, CO: Westview Press.