American Psychological Association

Developmental Psychology

© 2019 American Psychological Association
0012-1649/19/$12.00

http://dx.doi.org/10.1037/dev0000734

Children Use Targets' Facial Appearance to Guide and Predict Social Behavior

Tessa E. S. Charlesworth, Sa-kiera T. J. Hudson, Emily J. Cogsdill, Elizabeth S. Spelke, and Mahzarin R. Banaji
Harvard University

Humans possess a tendency to rapidly and consistently make character evaluations from mere facial appearance. Recent work shows that this tendency emerges surprisingly early: children as young as 3-years-old provide adult-like assessments of others on character attributes such as "nice," "strong," and "smart" based only on subtle variations in targets' face shape and physiognomy (i.e., latent face-traits). The present research examined the behavioral consequences of children's face-trait judgments by asking whether, and if so when in development, the appearance of face-traits also (a) shapes children's judgments *of* targets' behaviors and (b) guides children's behavior *toward* targets. Experiments 1 and 2 showed that, by 3 years of age, children used facial features in character evaluations but not in judgments of targets' behavior, whereas by 5 years of age, children reliably made both character and behavior judgments from face-traits. Age-related change in behavior judgments was also observed in children's own behaviors *toward* targets: Experiments 3 and 4 showed that, by age 5 (but not earlier), children were more likely to give gifts to targets with trustworthy and submissive-looking faces (Experiment 3) and showed concordance between their character evaluations and gift-giving behaviors (Experiment 4). These findings newly suggest that, although children may rapidly make character evaluations from face-trait appearance, predicting and performing social behaviors based on face-traits may require more developed and specific understanding of traits and their relationships to behaviors. Nevertheless, by kindergarten, even relatively arbitrary and subtle face-traits appear to have meaningful consequences in shaping children's social judgments and interactions.

Keywords: social–cognitive development, face perception, impression formation, trait inferences, behavior inferences

Supplemental materials: http://dx.doi.org/10.1037/dev0000734.supp

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Faces provide a rich source of information about others' identity and character, which subsequently has value for predicting others' behavior. In mere milliseconds, humans automatically detect gender, age, and race from static, two-dimensional (2D) face images based on features such as face shape, eye width, and skin-tone (Freeman & Johnson, 2016; Macrae & Bodenhausen, 2000). Not only do perceivers use such facial information to infer a person's social group membership, perceivers also use facial features to form quick evaluations of a person's general character (i.e., their overall positivity/negativity), as well as more specific impressions of a person's traits, such as their trustworthiness, dominance, and competence (Todorov, Pakrashi, & Oosterhof, 2009; Willis & Todorov, 2006).

Computational models of face-based judgments have made substantial progress in identifying the facial features that define such character evaluations and trait impressions. Specifically, impressions of trustworthiness largely correspond to the appearance of positive/negative emotions, impressions of dominance correspond, in part, to variations in chin shape and eyebrow width, and impressions of competence correspond, in part, to variations in eye width and eye size (Oosterhof & Todorov, 2008). These facial appearance cues are herein referred to as "face-traits" because they are physiognomic manipulations that are computationally modeled to evoke latent traits of trustworthiness, competence, or dominance.

Although the appearance of face-traits is correlated with other facial information, such as attractiveness and baby-facedness (Oosterhof & Todorov, 2008; Todorov, Dotsch, Porter, Oosterhof, & Falvello, 2013), face-traits have also been shown to provide additional explanatory value even when controlling for these other facial features (Duarte, Siegel, & Young, 2012; Graham, Harvey, & Puria, 2017; Oh, Buck, & Todorov, 2019). Indeed, over a decade of research has now shown that manipulating specific

Tessa E. S. Charlesworth, Sa-kiera T. J. Hudson, Emily J. Cogsdill, Elizabeth S. Spelke, and Mahzarin R. Banaji, Department of Psychology, Harvard University.

Emily J. Cogsdill is now in Private Practice at Austin, Texas.

This research was supported, in part, by awards from the E.J. Safra Center for the Study of Ethics and the Dean's Competitive Fund for Promising Scholarship to Mahzarin Banaji. All data, analyses, and stimuli are publicly available on OSF: https://osf.io/ukfzs/.

Correspondence concerning this article should be addressed to Tessa E. S. Charlesworth, Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, MA 02138. E-mail: tet371@g.harvard.edu

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

physiognomic face-traits correspondingly alters adults' impressions of faces (Todorov, 2017; Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015). The study of face-trait judgments therefore provides a window into unique processes of person perception.

Although face-traits are not necessarily accurate reflections of a target's true character (Olivola, Funk, & Todorov, 2014), adults' face-trait judgments have been shown to have real-world consequences in behavior, including predicting success in elections (Antonakis & Dalgas, 2009; Lawson, Lenz, Baker, & Myers, 2010; Todorov, Mandisodza, Goren, & Hall, 2005), rates of hiring and compensation (Graham et al., 2017), financial investment (Duarte et al., 2012; Rezlescu, Duchaine, Olivola, & Chater, 2012), and even the harshness of criminal sentencing (Blair, Judd, & Chapleau, 2004; Porter, ten Brinke, & Gustaw, 2010; Wilson & Rule, 2015). Importantly, across these outcomes, adults appear to use face-traits to both make judgments about the behaviors *of* targets (e.g., judging that a competent-looking face is more likely to win an election than an incompetent-looking face), and modify their own behaviors *toward* targets (e.g., giving a harsher criminal sentence to a dominant-looking face than a submissive-looking face). Thus, research from social and cognitive psychology indicates that the appearance of face-traits is used rapidly and consistently, even beyond other face information, in guiding adults' character and behavior judgments.

Emergence of Face-Trait Character Evaluations

Surprisingly, the capacity to make rapid and consistent character evaluations from face-traits emerges early in development and appears to require relatively less experience than other social cognitions (e.g., control of explicit prejudice emerges relatively late at around 10 years of age, Baron & Banaji, 2006; Raabe & Beelmann, 2011). Indeed, children as young as 3 years of age make congruent face-trait character evaluations, in that they are more likely to judge a competent-looking face (vs. incompetent-looking face) as "smart," a dominant-looking face (vs. submissive-looking face) as "strong," and a trustworthy-looking face (vs. untrustworthy-looking face) as "nice" (Cogsdill, Todorov, Spelke, & Banaji, 2014). By 5 years of age, children's generalized character evaluations from face-traits (i.e., their general impressions of targets along a positive/negative dimension) appear indistinguishable from the judgments of adults. Moreover, such early emergence of face-trait character evaluations is robust, occurring regardless of whether the targets are computer-generated faces, naturalistic child or adult faces, or even unfamiliar faces of another species (e.g., rhesus macaques, Cogsdill & Banaji, 2015).

The foundations for such robust face-trait evaluations has been suggested to appear as early as 7 months of age, when infants prefer to look at trustworthy-looking over untrustworthy-looking faces (Jessen & Grossmann, 2016). Notably, infants' preferences for trustworthy appearance may be related to similarly early emerging preferences for attractiveness (Langlois et al., 2000, 1987) and positive emotional expressions (Farroni, Menon, Rigato, & Johnson, 2007), implying that face-trait character evaluations may have their initial foundations in generalized evaluations of positivity/negativity. Indeed, preschoolers' character evaluations from other sources of information (e.g., past behaviors) show a generalized evaluation that gradually becomes differentiated into specific trait inferences across childhood (Brosseau-Liard & Birch, 2010; Cain, Heyman, & Walker, 1997). This transition from generalized evaluations to specific inferences is tied to a developing understanding of traits as specific, unique, and stable features that have consequences for predicting an individual's behaviors (Heyman, 2009). Analogously, children's face-trait character judgments may also transition from generalized evaluative responses in infancy to an understanding of face-traits as communicating more trait-specific information.

In these ways, developmental studies have begun to provide insights into the developmental precursors, use, and representation of latent face-traits in character evaluations. Yet previous research leaves open the question of whether, and if so when, children perceive such face-traits to have *behavioral* consequences. That is, children may use the appearance of trustworthiness, dominance, and competence to make character evaluations about general positivity and "niceness" (Cogsdill et al., 2014) but not yet understand the implications of face-traits in judging *others'* behaviors and in shaping their *own* behaviors toward others. Examining the emergence of face-trait behavior judgments, when interpreted alongside concurrent age-related changes in face-trait character evaluations, can inform our understanding of when, why, and how children and adults perceive face-traits as consequential information for social decisions.

Emergence of Face-Trait Behavior Judgments

As discussed above, numerous studies suggest that adults use face-traits when judging the likely behaviors *of* others as well as when performing behaviors *toward* others (for reviews see Olivola et al., 2014; Todorov, 2017). Developmental studies have only recently begun to explore whether and when children similarly use face-trait information to both judge and perform behavior. Children, from at least age 5, are more likely to select a face that has been prerated as competent-looking when selecting "the captain of the boat," implying that they use face-trait cues of competence in judgments of others' competence-relevant behavior (Antonakis & Dalgas, 2009). Furthermore, children, again from age 5, are more likely to perform trust behaviors, such as investing in, or believing the information of, faces that are prerated as trustworthy-looking or attractive (Bascandziev & Harris, 2016; Ewing, Caulfield, Read, & Rhodes, 2015).

Additional insights into the possible patterns of children's face-trait behavior judgments and performance comes from complementary research showing that children both predict and perform behaviors based on information about a target's previous behaviors (e.g., "behavior-to-behavior" inferences) especially in the domain of social learning (Sobel & Kushnir, 2013). For example, preschool-aged children predict that a previously accurate informant will also be accurate in the future and will therefore choose such an informant as a partner in a knowledge-related task (Fusaro, Corriveau, & Harris, 2011; Hermes et al., 2016; Hermes, Behne, & Rakoczy, 2015). Notably, such behavior-to-behavior inferences appear to rely on children's underlying trait reasoning. That is, past behaviors are interpreted to be diagnostic of an underlying trait, which, in turn, is used to judge the likelihood of future behavior (e.g., Chen, Corriveau, & Harris, 2016; Hermes et al., 2015). Similarly, behavior judgments from face-traits may also require proficiency in trait reasoning, such that children make a character evaluation from face-traits and then take that character evaluation as an input to judge the behaviors of targets as well as guide behaviors toward targets.

Together, the initial evidence of children's face-trait behavior judgments, as well as evidence from parallel research on behavior-to-behavior inferences, suggests that children's face-trait judgments may indeed have behavioral consequences, at least by age 5 and at least for specific face-traits of trustworthiness and competence. However, it remains to be seen whether the ability to judge and perform behaviors based on face-traits emerges *before* 5-years-old, alongside face-trait character evaluations, which emerge by at least 3-years-old (Cogsdill & Banaji, 2015; Cogsdill et al., 2014). Alternatively, face-trait behavior judgments may emerge after character evaluations because of later-developing understanding of targets' behavioral consistency and traits (Heyman, 2009; Rholes & Ruble, 1984). Indeed, children appear to possess only a tenuous understanding of traits and their behavioral consequences until at least 9 years of age (Alvarez, Ruble, & Bolger, 2001). Additionally, it is possible that even older children may need further experience into adulthood in order to track relationships between the relatively inaccurate and arbitrary face-traits (Olivola et al., 2014) and the types of behaviors performed toward these targets. Evidently, it is important to examine simultaneous age-related change in *both* character and behavior judgments over an extended age range (including children before age 5 as well as adults) in order to understand when and why face-traits gain behavioral significance.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

The Present Research

The present research was designed to examine whether, and if so when in development, children use face-trait information to both judge and perform consequential social behaviors. We examine children from 3- to 13-years-old, in line with previous studies investigating children's behavior and character judgments from past behavior (e.g., Alvarez et al., 2001; Stipek & Daniels, 1990), while also extending the age range of previous studies on children's face-trait behavior judgments (Antonakis & Dalgas, 2009; Ewing et al., 2015). Importantly, this age range captures other relevant developmental shifts in social cognition, including shifts from generalized to specific behavior inferences, occurring between 7 and 9 years of age (Alvarez et al., 2001), and shifts in trait reasoning (Heyman, 2009). Concurrent patterns of age-related change can help inform discussions of the possible mechanisms behind any observed change in face-trait judgments.

An additional advantage of the present research is the examination of face-trait judgments across multiple trait-relevant behaviors beyond leadership selection, investment, or selective learning. Including multiple behaviors, ranging from children judging "who can pick up a heavy box" to children performing prosocial gift-giving, offers a test of the robustness and generalizability of children's face-trait behavior judgments.

Furthermore, the current work examines judgments from multiple face-traits simultaneously, rather than examining trustworthiness or competence in isolation. Explicit social judgments are argued to be organized along fundamental axes of warmth (trustworthiness) and competence (Fiske, Cuddy, Glick, & Xu, 2002), while computational models have indicated that face-based judgments can be organized along dimensions of trustworthiness and dominance (Oosterhof & Todorov, 2008). As such, using three face-traits—trustworthiness, dominance, and competence—captures foundational dimensions of social evaluations and thereby facilitates claims about the extent and generalizability of children's face-trait judgments.

In sum, four experiments were conducted to address whether, and if so when, the appearance of latent face-traits shapes children's and adults' judgments about the behaviors *of* targets (Experiments 1 and 2) as well as guides children's and adults' own prosocial behaviors *toward* targets (Experiments 3 and 4). The research sheds new light on these questions by using: (a) an extended age range to assess patterns of age-related change; (b) concurrent examination of both character evaluations and behavior judgments to inform discussions of mechanisms; and (c) judgments for multiple face-traits (i.e., trustworthiness, competence, and dominance) and multiple behaviors to examine robustness across fundamental dimensions of social perception.

Experiment 1: Face-Traits Cues in Character Evaluations and Behavior Judgments

Experiment 1 aimed, first, to replicate previous findings of early emerging developmental consistency in generalized face-trait character evaluations of "niceness" and, second, to newly investigate whether and when face-traits also inform children's judgments of face-trait relevant behavior.

Method

Methods for all experiments were approved by the Committee on the Use of Human Subjects at Harvard University, under protocol IRB 21291 titled "Children's Facial Judgments." All data, stimuli and analysis scripts are available on the OSF at https://osf.io/ukfzs/.

Participants. A total of 99 children (M_{age} = 6.48 years [3.21, 10.67], SD = 1.92) participated in the experiment. No children were excluded from analyses. Approximately half of the child participants were identified by their parents as female (57%), and most were identified as White (47%), with the remainder identified as Asian (17%), Hispanic (9%), multiracial (9%), Black or African American (3%), or other/unknown (14%). In addition, 50 adults (M_{age} = 32.2 years [19, 66], SD = 10.85) participated online. Approximately half of the adult sample was female (52%), and the majority identified as White (72%), with the remainder identifying as Asian (12%), Hispanic (4%), multiracial (4%), Indian (2%), Black or African American (2%), or other/unknown (4%).

Post hoc power analyses based on 1,000 simulations conducted using the *simR* package (Green & Macleod, 2016) in the R computing environment (R Core Team, 2017) indicated that the sample provided power of 1.00 [.99, 1.00] (within machine precision) to detect the significant effect of age in children's responses.

Stimuli. Experiment 1 used computer-generated face stimuli from publicly available sets created in Face Gen 3.1 (Oosterhof & Todorov, 2008; Todorov et al., 2013; Todorov & Oosterhof, 2011). Three face-traits were used: faces that varied on the appearance of trustworthiness (trustworthy-untrustworthy), dominance (dominant-submissive), and competence (competent-incompetent). The set of faces were either 3 SDs above or below the average face on each trait dimension, making the faces strongly capture one of the three traits

(Figure 1, top row). Each trait dimension was represented by four faces for a total of 12 pairs of 24 faces.

Procedure. Children viewed pairs of faces on a laptop computer at a local children's museum and were guided through three practice trials using faces not included in the test trials. The practice trials acclimated children to the task and confirmed that they could clearly point to faces on the screen. All participants then completed one block of character evaluation questions and one block of behavior judgment questions, with order of block counterbalanced across participants. Each block had 12 trials for a total of 24 trials. For each trial, children viewed pairs of faces sequentially and answered character or behavior prompts by pointing to their chosen face on the screen. These pairs of faces were always the extreme ends of a single face-trait dimension: for example, a competent-looking face was always paired with an incompetent-looking face. At the conclusion of the experiment, children were offered their choice of a sticker and their adult family members were verbally debriefed.

Within the character evaluation block, participants saw 12 pairs of faces and were asked which face they thought was "mean" or "nice," capturing a general evaluation of positivity or negativity. The mean or nice question order was randomized for each participant. Within the behavior judgment block, participants saw 12 new pairs of faces (i.e., the same faces from the character evaluation block but in different pairings) and were asked which of the faces would perform a certain behavior (e.g., displaying strength) that was associated to the trait varied across the face pair (e.g., dominance; Figure 2). For example, when participants were shown a dominant-looking and a submissive-looking face side-by-side, they were asked "Which of these people can pick up really heavy things?" Each face-trait dimension was represented by two behaviors for a total of six behaviors that were repeated twice within the block.[1] For each participant, the three types of behavior trials proceeded in one of six possible fixed orders (e.g., trustworthiness-dominance-competence), which were counterbalanced across participants.

Adults on Amazon Mechanical Turk completed an online version of the experiment through www.socialsci.com. The procedure was nearly identical to that used for children with two minor differences. First, adults were not given any practice trials. Second, whereas exactly half of all trials in the children's version were configured to have anticipated responses on the left and right sides of the screen, the order and side in which faces appeared to adults was completely randomized.

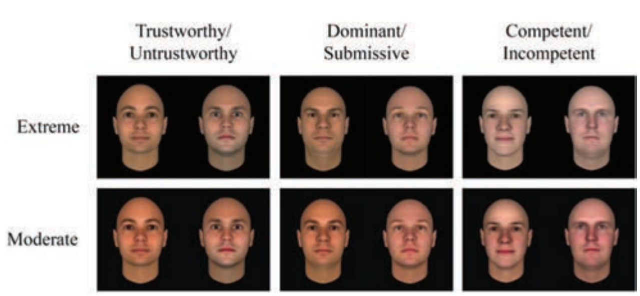

Figure 1. Face stimuli. Examples of computer-generated face stimuli portraying each face-trait in Experiments 1, 3, and 4 (+/− 3 *SD*, "Extreme" faces) and in Experiment 2 (+/− 1 *SD*, "Moderate" faces). All stimuli were retrieved from publicly available sets created in Face Gen 3.1 (see Todorov, Dotsch, Porter, Oosterhof, & Falvello, 2013; Todorov & Oosterhof, 2011). See the online article for the color version of this figure.

Trait	Behavior *("Which of these people _____")*
Trustworthiness	"helps other people when they are in trouble?" "likes to share their goods with other people?"
Dominance	"can pick up really heavy things?" "always decides which game to play?"
Competence	"knows how to sing a lot of different songs?" "can draw pictures that look just like real life?"

Figure 2. Behavior judgment stimuli. Trait-specific behavior prompts (Experiments 1 and 2).

Results and Discussion

Analytic strategy. The dependent variable of interest was the categorical "expected" or "not expected" response for each character evaluation or behavior judgment question, with the "expected" response coded as selecting the face that was consistent with each face-trait from previous ratings. Specifically, based on previous research (Cogsdill et al., 2014), an "expected" response in character evaluations was coded as selecting the trustworthy-, submissive-, or competent-looking face as "nice," or the untrustworthy-, dominant-, or incompetent-looking face as "mean." Similarly, an "expected" response in behavior judgments was coded as selecting the face that corresponded to the behavior (e.g., choosing the dominant-looking face when asked "Who can pick up really heavy things?"; see Figure 2).

The main research question is whether, and if so when, children make character and behavior judgments in line with these expected responses. Following basic tests to examine whether children and adults were overall more likely than chance to give the expected response, patterns of age-related change in the expected character and behavior judgments were investigated. In the latter models, the predictors of interest were judgment type (behavior vs. character), age, and the interaction between judgment type and age.

Because of the age gap between the oldest children (11-years-old) and youngest adults (19-years-old), as well as the different ranges in children (3–11 years) versus adults (19–60 years), entering age as continuous with the whole sample of children and adults would violate the assumptions of a truly continuous predictor and lead to spurious estimates. As such, age-related change was examined in two ways. First, to examine the rate of continuous developmental change from ages 3–11 for children's responses alone, and to ensure that the findings were not due to the choice of

[1] These behaviors were chosen based on their face validity to capture the trait dimensions conveyed by each of the face pairs. However, to confirm that these behaviors were indeed clearly related to the perceived traits of the faces, additional norming data were collected from a sample of 50 adult participants through Amazon Mechanical Turk. In all cases, the behaviors chosen to represent the perceived traits were significantly more likely than chance to be categorized as the expected trait (e.g., "helps other people" was categorized as trustworthy) than either of the other two traits, all P(behavior-trait matches) $>$ 74%, all *ps* $<$ .001. Indeed, the median proportion of behavior-trait matches was 96% (see online supplementary materials), thereby confirming that these behaviors accurately captured the expected trait dimensions.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

specific age groups, a model was fit with age as a continuous variable in children's responses (Figure 3a).

Second, to compare responses of both children and adults while accounting for the age gap and range differences, age was coded using discrete categories: children ages 3–4 ($N = 24$), 5–6 ($N = 37$), 7–11 ($N = 38$), and adults ($N = 50$; Figure 3b). These age groups were chosen given previous research suggesting shifts in the stability and specificity of trait understanding between preschoolers (3- to 4-years-old) and kindergarteners (5- to 6-years-old; Brosseau-Liard & Birch, 2010), as well as yet more complete trait understanding developing around 7 years of age (Heyman, 2009). Results across the continuous-age and discrete-age models provided similar conclusions. Thus, for simplicity, the full results from the discrete-age model are presented in the online supplementary materials with highlights noted in the discussion below.

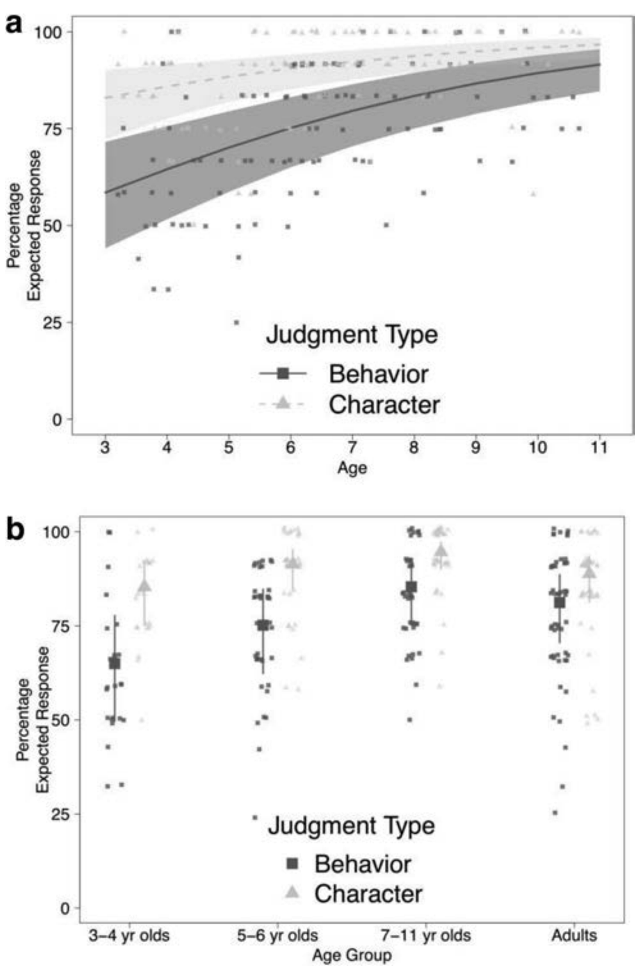

Figure 3. *a.* Effect of age in children's character evaluations and behavior judgments (Experiment 1). Confidence intervals and trajectories are predicted values from the interaction model of age (continuous) and judgment type, with random effects of subject and stimuli; raw values are plotted for all participants with jitter. *b.* Effect of age in children's and adult's character evaluations and behavior judgments (Experiment 1). Confidence intervals are predicted values from the interaction model of age group and judgment type, with random effects of subject and stimuli; raw values are plotted for all participants with jitter.

Results for all experiments were analyzed using generalized linear mixed-effects models (GLMM; a logistic regression with random effects) using the *lme4* package in R (Bates, Mächler, Bolker, & Walker, 2015). GLMM allows modeling of trial-level data (rather than aggregated data as in an ANOVA) and can therefore account for variations among stimuli using random intercepts. A random intercept of subject was entered to account for trial-level dependencies from repeated trials. Additionally, a random intercept of face stimuli (i.e., the specific face of the 24 possible faces used) nested within face-trait (i.e., whether that face varied on competence, trustworthiness, or dominance) was entered to account for random variation across stimuli. The random intercepts were entered successively and the models were compared using likelihood ratio tests to assess whether the addition of each parameter resulted in a significantly better fit to the data.

Following random effects specification, the fixed effects of age, judgment type (character or behavior), and their interaction were entered successively as predictors. Additionally, all models included covariates of respondent gender and race, which were grand-mean-centered such that the intercept could be interpreted as reflecting the "average" gender or race (Kreft & de Leeuw, 1998). Again, successive fixed effects models were compared using likelihood ratio tests to select the best-fitting and most parsimonious final model. Results for the random effects and fixed effects model comparisons are reported below.

Overall effects. To first examine the overall rate of expected responses, one-sample t tests were performed to compare all participants' responses with chance (50%). Overall, child and adult participants were significantly more likely than chance to select the trustworthy-, competent-, or submissive-looking face as "nice," or the untrustworthy-, incompetent-, or dominant-looking face as "mean," thus making the expected face-trait character evaluations ($M = 88\%$ [86%, 90%], $SD = 12\%$), $t(148) = 37.43$, $p < .001$, $d = 3.07$. Child and adult participants were also significantly more likely than chance to provide the expected behavior judgment based on face-traits (e.g., picking the dominant-looking face as the face that would "pick up really heavy things," $M = 75\%$ [73%, 78%], $SD = 17\%$), $t(148) = 17.71$, $p < .001$, $d = 1.45$. Although participants were more likely than chance to provide the expected response for both judgments, they remained more likely to select the expected face in character evaluations than in behavior judgments, $t(265.59) = 7.04$, $p < .001$, $d = 0.82$.

Interaction of age and judgment type in children's face-trait judgments. For modeling children's judgments alone (as described above), the model with two random intercepts of subject and stimuli significantly improved model fit above the previous model with only a random intercept of subject, $\chi^2(2) = 48.22$, $p < .001$. The model with two fixed effects for age (continuous) and judgment type significantly improved model fit beyond the previous model with a single main effect of judgment type, $\chi^2(1) = 26.63$, $p < .001$. However, the fixed effect model with an interaction of age and judgment type did not provide significantly better fit above the previous model with two main effects, $\chi^2(1) = 0.20$, $p = .65$.

This suggests that children improved with age at similar rates in both character and behavior judgments, such that with each year the odds that children would select the expected face increased by 1.28, $OR = 1.28$, $z = 5.50$, $p < .001$. Indeed, collapsing across both character and behavior judgments, chil-

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

dren at the youngest ages had a predicted rate of selecting the expected face of $M_{Predicted}$ = 72% [59%, 82%], while children at the oldest ages had progressed to near-ceiling rates of selecting the expected face, $M_{Predicted}$ = 93% [89%, 96%].

Despite the similar rates of change, however, the starting point for children's character and behavior judgments differed, such that the overall odds of children selecting the expected face were more than three times greater in character evaluations ($M_{Predicted}$ = 92% [87%, 95%]) than in behavior judgments ($M_{Predicted}$ = 77% [68%, 85%]), $OR = 3.16$, $z = 9.63$, $p < .001$. These conclusions were replicated using age as a discrete category and including adult responses (see Figure 3b and online supplementary materials). For additional interpretation regarding the differences in starting rates of expected responses, separate models were fit for character and behavior judgments using age (continuous) as the single predictor and the same random intercepts of subject and stimuli.

Effect of age in children's character evaluations. Children at the youngest ages (i.e., at the intercept) were significantly more likely than chance to select the expected face in character evaluations ($M_{Predicted}$ = 84% [73%, 91%]), $OR = 5.38$, $z = 5.03$, $p <$ .001. With each year of age, the odds that children selected the expected face in character evaluations significantly increased by 1.28, $OR = 1.28$, $z = 3.71$, $p < .001$. Thus, by the oldest ages, children were selecting the expected face at a near-perfect rate of $M_{Predicted}$ = 97% [93%, 99%].

Effect of age in children's behavior judgments. With age, children also became significantly more likely to make expected responses in behavior judgments. However, the intercept indicated that the behavior judgments of the youngest children were not significantly different from chance ($M_{Predicted}$ = 60% [44%, 74%]), $OR = 1.49$, $z = 1.23$, $p = .21$. Nevertheless, with each year the odds for children to give an expected response in behavior judgments significantly increased by 1.29, $OR = 1.29$, $z = 5.25$, $p < .001$, paralleling the rate of change in children's character evaluations. At the oldest ages, children were selecting the expected face in behavior judgments at $M_{Predicted}$ = 90% [82%, 95%], nearly converging with the oldest children's responses in character evaluations. Indeed, inspection of the predicted values shows a 24-percentage point gap between character and behavior judgments for the youngest children that narrows to seven percentage points for the oldest children.

These results newly show that consistent character and behavior judgments from face-traits have different moments of emergence, passing chance by at least 3 years of age for character evaluations but by 5 years of age for behavior judgments. Indeed, it appears that achieving proficiency in face-trait behavior judgments is up to 4 years behind face-trait character evaluations: Three- to 4-year-olds make the expected character evaluation 85% of the time, whereas it takes until 7–11 years of age to make the expected behavior judgment at this same rate. Thus, experience in making face-trait character evaluations may be required as an input to the development of face-trait behavior judgments. Such findings align with previous research showing that children's behavior predictions from a target's past behavior also rely on trait reasoning (Chen et al., 2016; Hermes et al., 2015), and thereby implies that both behavior-to-behavior and face-to-behavior judgments may require developing an understanding of traits as stable features of individuals with consequences for future behavior.

Notably, the results also show that behavior and character judgments follow parallel rates of continuous change throughout childhood (Figure 3a) and even into adulthood (Figure 3b), and that neither judgment type shows a single discontinuous development shift. Instead, each age group appears successively more consistent than the last. The perception of face-traits as consequential for both character evaluations and behavior judgments may therefore need continued experience throughout childhood and adulthood to track relationships between faces, character, and behaviors, a point we return to in the General Discussion section.

Experiment 2: Subtle Face-Trait Cues in Character Evaluations and Behavior Judgments

Experiment 1 demonstrated that children as young as 5-years-old can use face-trait information to generate consequential judgments about the relative likelihood of others' behavior. However, the faces used in Experiment 1 were selected from the extreme ends of the face-traits and were therefore unusually strong exemplars. Because the most common faces encountered in the natural world vary more subtly in signaling traits, a test with moderately distinctive faces would provide more convincing evidence that character and behavior judgments are likely to occur in children's and adults' everyday interactions. Given that children have been found to make character evaluations from subtle, naturalistic, and novel faces to similar degrees as adults (Cogsdill & Banaji, 2015), Experiment 2 examines whether children can also make behavior judgments from subtle faces to similar degrees as adults. Experiment 2 also provides an internal replication of the new findings from Experiment 1 and gives insights into the generalizability and limits on the developmental emergence of face-trait behavior judgments.

Method

Participants. A total of 107 children (M_{age} = 6.51 years [3.24, 10.90], $SD = 1.92$) participated at a local children's museum. No data were excluded from analyses. The majority of participants were identified by their parents as female (55%) and White (63%), with the remainder identified as Asian (10%), multiracial (6%), Black or African American (2%), Hispanic (1%), or other/unknown (19%). A separate group of 50 adults (M_{age} = 34.54 years [19, 57], $SD = 11.06$) participated online. This sample was predominantly female (60%) and White (76%), with the remainder identifying as Hispanic (8%), Black or African American (8%), Asian (4%), multiracial (2%), or other/unknown (2%). Again, post hoc power analyses using 1,000 simulations from the final best-fitting fixed-effects model indicated power of 1.00 [0.99, 1.00] (within machine precision) to detect the significant effect of age in children's judgments.

Stimuli and procedure. The experimental procedure was identical to Experiment 1, except that the face stimuli were moderate exemplars of the traits (±1 *SD* from average; Figure 1 "Moderate faces"). As in Experiment 1, the face stimuli were accessed from the publicly available database produced in FaceGen 3.1 (Todorov et al., 2013; Todorov & Oosterhof, 2011).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Results and Discussion

Analytic strategy. Data analyses followed the same strategy as Experiment 1, using a GLMM with the categorical dependent variable "expected" or "not expected." The models successively entered random effects of subject and stimuli, fixed effects of age, judgment type, and their interaction, as well as grand-mean centered covariates of gender and race. All model comparisons are reported below. As before, due to age gaps and differences in age ranges between children and adults, two models were fit. For models with children only, age was continuous, whereas for models with both children and adults, age was entered as a discrete variable with the same age categories as Experiment 1, providing the following sample sizes: ages 3–4 ($N = 25$), 5–6 ($N = 41$), 7–11 ($N = 41$), and adults ($N = 50$). Results from models with continuous-age and discrete-age provided similar conclusions, therefore the model using discrete-age is reported in the online supplementary materials, with highlights noted below.

Overall effects. As in Experiment 1, one-sample *t* tests indicated that child and adult participants were significantly more likely than chance to make the expected character evaluation, selecting the trustworthy-, competent-, or submissive-looking face as "nice" ($M = 76\%$ [74%, 79%], $SD = 15\%$), $t(156) = 22.51$, $p < .001$, $d = 1.80$. Children and adults were also significantly more likely than chance to make the expected face-consistent behavior judgment ($M = 68\%$ [65%, 70%], $SD = 16\%$), $t(156) = 13.49$, $p < .001$, $d = 1.08$. Additionally, participants were significantly more likely to make expected responses in character evaluations than in behavior judgments, $t(308.15) = 4.95$, $p < .001$, $d = 0.56$.

Interaction of age and judgment type in children's face-trait judgments. For modeling age-related change in children's judgments the model with two random intercepts of subject and stimuli significantly improved model fit above the previous model with only a random intercept of subject, $\chi^2(2) = 22.72$, $p < .001$. As in Experiment 1, the model with two fixed effects for age and judgment type significantly improved model fit beyond the previous model with a single main effect of judgment type, $\chi^2(1) = 18.45$, $p < .001$. However, the interaction model did not provide significantly better fit above the previous model with two main effects, $\chi^2(1) = 0.01$, $p = .92$.

This indicates that, as for Experiment 1, children changed at similar rates in both character evaluations and behavior judgments (see Figure 4): With each year the odds of selecting the expected face increased by 1.15, $OR = 1.15$, $z = 4.45$, $p < .001$. Descriptively, across character and behavior judgments children at the youngest ages had a predicted rate of selecting the expected face of $M_{Predicted} = 61\%$ [54%, 68%], while children at the oldest ages had progressed to $M_{Predicted} = 81\%$ [76%, 85%].

Additionally, the overall odds of making an expected response were 1.64 times greater for children's character evaluations ($M_{Predicted} = 77\%$ [73%, 81%]) than for children's behavior judgments ($M_{Predicted} = 67\%$ [62%, 71%]), $OR = 1.66$, $z = 5.64$, $p < .001$. For additional interpretation on these different starting points, separate models were fit for character and behavior judgments using age (continuous) as the single predictor and the same random intercepts of subject and stimuli.

Effect of age in children's character evaluations. As in Experiment 1, with age, children became significantly more likely to make the expected face-trait character evaluations. Children at the youngest ages were already performing significantly better than chance ($M_{Predicted} = 68\%$ [60%, 76%]), $OR = 2.16$, $z = 4.30$, $p < .001$, and with each year, the odds of making an expected response increased by 1.15, $OR = 1.15$, $z = 3.20$, $p = .001$. By the oldest ages children were making expected character evaluations at $M_{Predicted} = 87\%$ [80%, 91%].

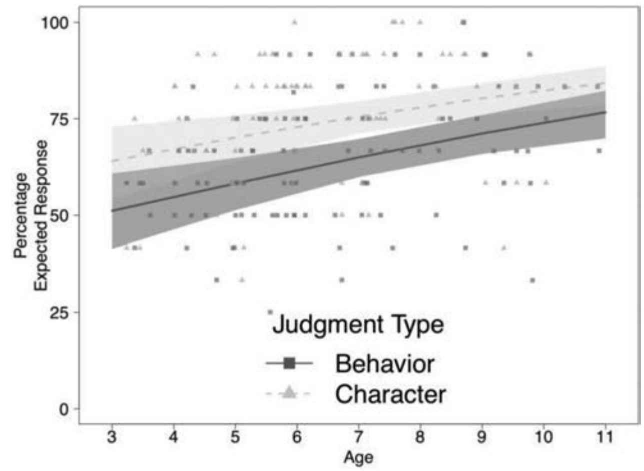

Figure 4. Effect of age in children's character evaluations and behavior judgments from subtle stimuli (Experiment 2). Confidence intervals and trajectories are predicted values from interaction model of age and judgment type, with random effects of subject and face-trait type; raw values are plotted for all participants with jitter.

Effect of age in children's behavior judgments. With age, children also became significantly more likely to make the expected responses in face-trait behavior judgments. However, as in Experiment 1, the intercept indicated that the behavior judgments of the youngest children were not significantly different from chance ($M_{Predicted} = 55\%$ [46%, 64%]), $OR = 1.22$, $z = 1.11$, $p = .27$. Nevertheless, with each year, the odds of making an expected response increased by 1.16, $OR = 1.16$, $z = 3.99$, $p < .001$, such that, by the oldest ages, children were making expected behavior judgments at $M_{Predicted} = 80\%$ [71%, 86%]. Comparing the model-predicted percentages shows that the difference between rates of expected character and behavior judgments converges across development: Whereas there was a difference of 13 percentage points for the youngest ages, there was a difference of only seven percentage points for the oldest ages.

The replication of findings across Experiments 1 and 2 suggests that children's and adults' capacity to make consistent character evaluations and behavior judgments from face-traits reflects a social–cognitive tendency that is robust to variations in the strength of stimuli (i.e., weak vs. strong exemplars of face-traits). The results again show that character and behavior judgments have different moments of developmental emergence but nevertheless follow similar rates of age-related change and ultimately converge in adulthood. Thus, although face-traits are not necessarily accurate cues to true character (Olivola et al., 2014; Olivola & Todorov, 2010) and, in the case of Experiment 2 are also relatively subtle cues, children from as early as kindergarten appear to use face-traits in meaningful judgments of others' behaviors.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Experiments 3: Face-Trait Cues in Performance of Prosocial Behaviors

Given that children and adults use face-trait information not only to make global character evaluations but also to judge the relative likelihood of trait-congruent behaviors (Experiments 1 and 2), it is possible that children may even perceive face-trait information as sufficiently meaningful to guide their own behavior toward targets. That is, it is possible that children will also *act* on the information evoked by face-traits. While previous studies of adults and children from age 5 have shown that perceived facial trustworthiness guides children's behaviors in trust decisions (Ewing et al., 2015; Rezlescu et al., 2012), Experiment 3 extends these findings in three ways. First, rather than focusing on specific trust decisions as the behavioral outcome, the current experiments assess generalized prosocial behaviors using a variant of the gift-giving paradigm (Kinzler, Dupoux, & Spelke, 2012). Assessing more generalized behaviors provides closer alignment with the generalized character evaluations of "niceness" used in Experiments 1 and 2. This helps address a potential concern regarding whether the later-emergence of behavioral judgments was due to the particular difficulty of making specific versus general judgments, rather than the difference between behavior versus character judgments per se.

Second, rather than focusing on a single trait of trustworthiness, as in previous work, Experiment 3 includes faces from multiple traits (trustworthiness and dominance), thereby illuminating the robustness of face-trait information in guiding prosocial behaviors. Additionally, comparisons across traits of trustworthiness and dominance can provide initial insights into whether children are giving gifts based on generalized halo effects (e.g., Brosseau-Liard & Birch, 2010; Cain et al., 1997) or based on specific expectations of reciprocity (e.g., McAuliffe, Raihani, & Dunham, 2017). That is, if children are giving gifts based primarily on generalized character evaluations, then they should give to faces that are typically perceived as "nicer" (i.e., submissive and trustworthy faces). In contrast, if children are giving gifts based on specific expectations of reciprocity, then they should give to faces that may be perceived to have greater control over resources and decisions (i.e., dominant and trustworthy faces). Supplemental analyses are reported to test these predictions across traits and provide insights into the possible explanations for children's face-trait behavior performance.

Third, a wide age range (ages 3–13) is used to inform discussions of the patterns and causes of age-related change in face-trait behavior performance. As aforementioned, this extended age range encompasses relevant concurrent age-related change in behavior and trait reasoning (Heyman, 2009) and prosocial norm understanding (McAuliffe et al., 2017). Such concurrent trajectories can facilitate interpretations of the sources of age-related change in children's prosocial behaviors based on face-traits. In these ways, the experiments provide new insights into when and how children may use face-trait information to shape their own behaviors.

Method

Participants. A total of 100 children (M_{age} = 7.19 [3.22, 13.17], SD = 2.37) participated at a local children's museum. No data were excluded from analysis. The majority of the sample was identified by their parents as female (72%) and White (83%), with the remainder identified as Asian (7%), Hispanic (3%), Black (2%), or other/unknown (5%). Post hoc power analyses using 1,000 simulations from the best-fitting model indicated adequate power of 0.75 [0.73, 0.78] to detect the significant effect of age in children's judgments.

Stimuli and procedure. Children viewed pairs of faces that were extreme exemplars of face-traits of trustworthiness or dominance (as in Experiment 1), generated in FaceGen 3.1 (Todorov et al., 2013; Todorov & Oosterhof, 2011). Extreme exemplars were used to elicit the strongest reaction, as this served as an initial proof-of-concept study to examine whether children would use face-traits to guide their own behaviors. As in the design of Kinzler et al. (2012), for each trial children viewed face pairs in a binder and were given a laminated picture of a desirable item, such as a cookie or chocolate bar (see Figure 5). Children were then asked to give the gift to one of the faces. For example, children were told: "*This is Edgar, and this is Martin. If you had only one cookie, which person would you give it to? Edgar or Martin?*" The names of characters varied randomly across trials. Children responded by sticking the gift to the Velcro below their chosen face. Two practice trials used dogs and cats as targets, with a steak and a toy mouse as sample "gifts," to ensure that children understood the procedure. Children completed eight test trials, with four trustworthy-untrustworthy and four dominant-submissive face pairs. Four predetermined randomized orders were used, counterbalancing the side of the trustworthy- or submissive-looking (i.e., the "expected") face across orders.

Results and Discussion

Analytic strategy. As before, a GLMM was used to assess the effect of age on children's prosocial behaviors from face-trait information. The dependent variable was the categorical "expected" or "not expected" response, with expected responses coded as children giving gifts to the trustworthy- or submissive-looking faces. Random effects of subject, stimuli, and order were entered successively. In addition to grand-mean centered covariates of gender and race, the single fixed effect of age was entered as a continuous variable because only children were included in this experiment.

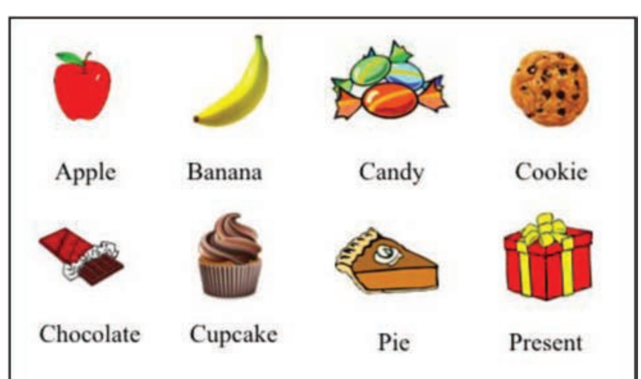

Figure 5. Gift stimuli used in Experiments 3 (all items) and 4 (top four items). See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Overall effects. Children were overall significantly more likely than chance to give their gift to the expected trustworthy- or submissive-looking face (M = 68% [64%, 73%], SD = 23%), $t(99) = 8.11$, $p < .001$, $d = 0.81$.

Effect of age in children's face-trait prosocial behaviors. The model with two random intercepts of subject and stimuli did not significantly improve model fit above the previous model with a single random intercept for subject, $\chi^2(2) = 0$, $p > .99$. The random effect of stimuli was therefore excluded to ensure model parsimony. The single fixed effect of age significantly improved model fit above the previous model that included only covariates of gender and race, $\chi^2(1) = 6.21$, $p = .01$.

The intercept for the fixed effect of age indicated that the youngest children (3-year-olds) were not significantly more likely than chance to give gifts to the trustworthy- or submissive-looking faces ($M_{Predicted}$ = 60% [49%, 69%]), $OR = 1.48$, $z = 1.78$, $p = .08$. However, with each year, the odds of performing an expected face-consistent behavior increased by 1.13, $OR = 1.13$, $z = 2.51$, $p = .01$, such that by the oldest ages children were giving gifts to the expected faces at a consistent rate of $M_{Predicted}$ = 83% [73%, 90%] (see Figure 6).

To better understand the mechanisms behind children's face-trait behavioral performance, a supplemental analysis was conducted using the fixed-effect interaction between age and face-trait to compare responses across the face-traits of trustworthiness and dominance. As discussed above, a significant interaction or main effect of face-trait would be predicted if children were giving gifts based on specific reciprocity expectations because they would give faces to the expected face only on trustworthy trials (i.e., the "nice" trustworthy face), but to the *unexpected* face on dominance trials (i.e., the "mean" dominant face). In contrast, neither an interaction nor a main effect of face-trait would be expected if children were giving gifts based on generalized halo effects because they would be giving gifts to the expected (i.e., "nice") face on both trustworthy and dominance trials. In line with the latter possibility, neither a significant interaction of age and face-trait, $OR = 0.99$, $z = -0.17$, $p = .87$, nor a main effect of trait, $OR = 1.10$, $z = 0.31$, $p = .75$, emerged. Full results of this model are described in the online supplementary materials. These exploratory analyses imply that the reasons for children's face-trait prosocial behaviors are more likely due to generalized evaluations and preferences rather than to specific expectations about reciprocity or resources.

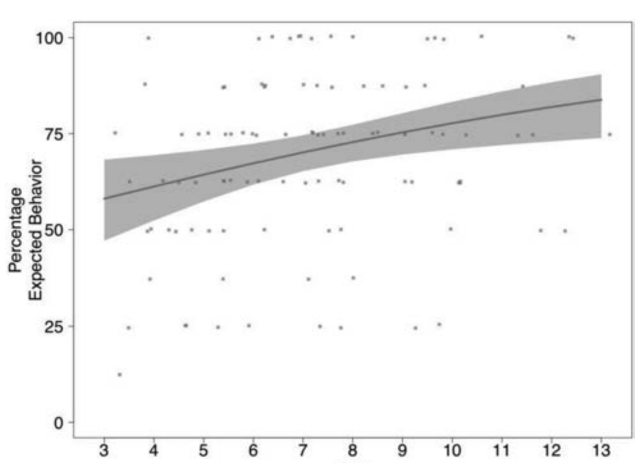

Figure 6. Effect of age in children's prosocial behaviors from face-traits (Experiment 3). Confidence intervals and trajectories are predicted values from the model with the fixed effect of age and random intercept of subject; raw values are plotted for all participants with jitter.

More generally, the results of Experiment 3 demonstrate that children not only judge the relative likelihood of trait-consistent behaviors from face-traits (Experiments 1 and 2) but also use face-trait information to guide their own prosocial behaviors. The act of giving gifts to "nicer"-looking faces, whether due to the appearance of trustworthiness or submissiveness, appears to emerge around age 5, but not earlier. The consistency of face-trait behavioral performance subsequently increases linearly over development, with each age group appearing successively more consistent than the last. In addition, supplemental exploratory analyses suggest that children's face-trait gift-giving was likely guided by generalized preferences or "halo effects" (Cain et al., 1997) rather than specific or strategic expectations of reciprocity. Such findings may be taken to suggest that, at least in early and middle childhood, face-trait behavior performance is based on general evaluations of good/bad that may gradually become more trait-specific. Future research exploring this possibility would benefit from including even older ages and including more face-traits and more behaviors (e.g., selecting partners for an athletic game or a knowledge task).

Overall, Experiment 3 newly suggests that, from early ages, generalized prosocial behavior is guided not only by evolutionarily meaningful social category membership (Kinzler et al., 2012), but also by relatively arbitrary cues of latent facial appearance of trustworthiness and dominance. In other words, even before entering formal schooling, children appear to commit face-based biases (Olivola et al., 2014) in their prosocial behaviors.

Experiment 4: Concordance of Character Evaluations and Prosocial Behaviors

The results presented thus far suggest that children's use of face-trait cues extends beyond character evaluations and has meaningful consequences for both behavioral judgments and performance. These findings also reveal developmental change: although face-trait character evaluations are robust by 3 years of age, consistent behavior judgments and performance from face-trait information appear only after 5 years of age. The later emergence of these behavior judgments and performance suggests that proficiency in face-trait character evaluations and underlying trait reasoning may be required as inputs to understanding the role of face-traits in judging and performing behaviors.

However, the experiments described thus far are limited in that they have measured character and behavior judgments for independent, nonoverlapping pairs of faces and, as such, cannot directly test the *concordance* between judgments. This leaves open the possibility that any observed convergence in the rates of expected responses for character and behavior judgments (as observed among older children and adults from Experiments 1 and 2) emerges only at the aggregate level across participants and face pairs. Convergence may not necessarily be present at the level of the individual participant and individual face pairs. This distinction is analogous to the statistical difference between alignment in

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

means (i.e., aggregate concordance) and alignment in correlations (i.e., individual concordance). The latter is particularly helpful when making inferences about the relationship of one variable (in this case, character evaluations) to a second variable of interest (in this case, behavior performance). Experiment 4 therefore provides an initial test of the individual-level concordance between face-trait character and behavior performance, to identify whether, and if so when, these types of judgments reveal alignment.

Method

Participants. A total of 43 children (M_{age} = 7.25 [4.02, 11.81], $SD = 2.22$) participated at a local children's museum. No data were excluded from analysis. The majority of children were identified by their parents as female (77%) and White (60%), with the remainder identified as Asian (14%), Black (2%) or other/unknown (23%). The sample for Experiment 4 was smaller than in previous experiments due to experimenter turnover. As such, the analysis of the significant effect of age in Experiment 4 may best be interpreted with caution, as it had a simulated post hoc power of 0.43 [0.40, 0.46] to detect the main effect of age.

Nevertheless, the same post hoc simulation-based power analysis approach indicated power of 0.55 [0.52, 0.58] to detect an intercept that was significantly different from chance, that is, to detect whether the youngest ages had concordance significantly above-chance. Additionally, the sample had a power of 0.98 to detect the overall effect size of children's judgments being more concordant than chance in a one-sample *t* test. Perhaps most convincingly, a meta-analysis of the odds ratio effects across all experiments (reported below) indicated that the results of Experiment 4 had a similar effect size to the meta-analytic estimate for the main effect of age. Thus, taken together, these arguments suggest that the effects from Experiment 4 may be interpreted as representative results that can provide insights into the emergent pattern of concordance between character and behavior judgments.

Stimuli and procedure. Face stimuli were identical to Experiment 1, using extreme exemplars of faces generated in FaceGen 3.1 (Todorov et al., 2013; Todorov & Oosterhof, 2011). All participants completed both a character evaluation task (as in Experiments 1 and 2) and a gift-giving task (as in Experiment 3). To accommodate two rounds of experimentation, only faces varying in trustworthiness and only the first four "gifts" were used (i.e., apple, banana, cookie, and cupcake; see Figure 3). Following three familiarization trials, children completed eight test trials of character evaluations (i.e., "mean" vs. "nice") for trustworthy-untrustworthy face pairs presented on a computer. Children then completed four test trials of gift-giving to a subset of these same face pairs presented in a binder. A subset of the initial character evaluation pairs was used to obscure the overlap between the two rounds of the game; no children spontaneously verbalized that they noticed the reappearance of the face pairs.

Results and Discussion

Analytic strategy. A GLMM was used to assess the effect of age on the concordance between children's face-trait behavior performance and character evaluations. The dependent variable was the categorical variable indicating whether children's behaviors performance and character evaluations showed a "match" or "no match." Random intercepts were subject, stimuli, and order, and the fixed effects were age (continuous) as well as gender and race entered as grand-mean centered covariates. Results of model comparisons are presented below.

Overall effects. One-sample *t* tests indicated that children were overall significantly more likely than chance to select the expected trustworthy-looking face as "nice" in character evaluations (M = 85% [79%, 91%], SD = 20%), $t(42) = 11.35$, $p < .001$, $d = 1.73$, as well as to give their gift to the expected trustworthy-looking face (M = 72% [62%, 82%], SD = 34%), $t(42) = 4.30$, $p < .001$, $d = 0.66$. Additionally, because children gave both character and behavior responses to the same faces it was possible to examine the concordance (i.e., "match") between their character evaluations and behavior performance. Children were overall significantly more likely than chance to make concordant character and behavior responses (M = 75% [66%, 84%], SD = 28%), $t(42) = 5.78$, $p < .001$, $d = 0.88$.

Effect of age on concordance of character evaluations and prosocial behaviors. Neither the model with three random intercepts of stimuli, order, and subject, nor the model with two random intercepts of stimuli and subject significantly improved model fit beyond the base model with a single random intercept for subject (all χ^2 s = 0, all *p*s > .99). Thus, only the random effect of subject was included to ensure model parsimony. The fixed effect of age did not significantly improve model fit above the model with only covariates of gender and age, $\chi^2(1) = 2.07$, $p = .15$. However, as aforementioned, the nonsignificant effect of age was likely due to low power, because the effect size for age from Experiment 4 aligned with the significant meta-analytic estimate for age. As such, the model with the single fixed effect of age is reported below.

The youngest children, at the intercept, were not significantly more likely than chance to show concordance between their character evaluations and behavior performance ($M_{Predicted}$ = 70% [47%, 86%]), $OR = 2.31$, $z = 1.70$, $p = .09$. However, the odds of showing concordance between character and behavior responses increased (albeit not significantly) by 1.21 every year, $OR = 1.21$, $z = 1.43$, $p = .15$. By the oldest ages, children were more likely than chance to make concordant judgments, with a predicted concordance of $M_{Predicted}$ = 91% [71%; 98%] (see Figure 7).

Descriptively, the results imply a gradual emergence across middle childhood in the concordance of children's character and behavior judgments from face-trait information. This reinforces the interpretation of results from Experiments 1 and 2 in suggesting that age-related change in children's face-trait behavior judgments may require earlier proficiency in face-trait character evaluations, until capacity in both social judgments eventually converges in later childhood and adulthood. Notably, this convergence has now been shown in both aggregate (Experiments 1 and 2) and individual-level (Experiment 4) experimental designs. Moreover, these findings show that, at least by kindergarten, children's judgments from face-traits have early emerging and meaningful behavioral consequences, even for static 2D images with relatively subtle and arbitrary variations (Olivola et al., 2014).

Meta-analysis of age-related change in children's face-trait character and behavior judgments. A primary focus of all four experiments was to examine patterns of age-related change in children's judgments from face-trait information, including generalized character evaluations, specific behavior judgments, and

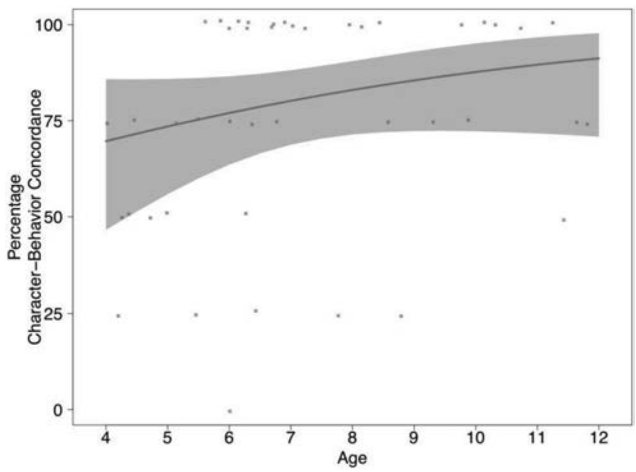

Figure 7. Effect of age in children's concordance of character evaluations and prosocial gift-giving behavior (Experiment 4). Confidence intervals and trajectories are predicted values from the model with the fixed effect of age and random effect of subject; raw values are plotted for all participants with jitter.

generalized prosocial behavior performance. Therefore, to provide a summarized estimate of the effect size and rate of age-related change a within-study meta-analysis was performed using the *metafor* package in R (Viechtbauer, 2010) using the odds ratio effect sizes of the effect of age across all experiments. The meta-analytic estimate indicated a significant effect size for age-related change in children's face-trait judgments, $OR = 1.24$ [1.15, 1.33], $z = 27.74$, $p < .001$. This implies that, across all experiments, the odds of making an expected or concordant judgment from face-trait information increased significantly by 1.24 with each year of age, suggesting meaningful age-related increases in the consistency of face-trait character evaluations, behavior judgments, and behavior performance.

General Discussion

Social perceivers provide rapid and consistent judgments of others' character on the basis of subtle variations in facial physiognomy and shape (Freeman & Johnson, 2016; Todorov et al., 2015). These character judgments are given reliably from surprisingly early ages (Cogsdill & Banaji, 2015; Cogsdill et al., 2014; Jessen & Grossmann, 2016). The current experiments newly show that such early emerging judgments from latent face-trait cues are not isolated to character evaluations, but rather, extend to both judgments and performance of behaviors. Specifically, children judged that a face with the latent appearance of trustworthiness would be more likely than an untrustworthy-looking face to perform trustworthy behaviors (e.g., share their goods), a face with the latent appearance of dominance would be more likely than a submissive-looking face to perform dominant behaviors (e.g., pick up heavy things), and a face with the latent appearance of competence would be more likely than an incompetent-looking face to perform competent behaviors (e.g., draw the best pictures).

Moreover, children were even found to use face-traits in guiding their own gift-giving behaviors in ways consistent with a target's face-trait appearance, preferring to give gifts to trustworthy- and submissive-looking faces over untrustworthy- or dominant-looking faces. Finally, these prosocial behaviors were found to show gradually emerging individual-level concordance with children's character evaluations, suggesting a developing understanding of the overlapping implications of face-trait appearance for impressions of both character and behavior.

Emergence of Face-Trait Character Evaluations, Behavior Judgments, and Behavior Performance

The capacity to make behavioral judgments and performance from face-traits has previously been demonstrated in children as young as 5-years-old for specific traits of competence (Antonakis & Dalgas, 2009) and trustworthiness (Ewing et al., 2015). However, until the present research, the earlier developmental trajectory of these behavior judgments remained unexplored, despite research documenting earlier face-trait character evaluations (Cogsdill et al., 2014). Replicating previous work, the current experiments show that face-trait character evaluations are consistently above-chance from 3 years of age. In contrast, face-trait behavioral judgments and performance remain at-chance for 3-year-olds and only reach high consistency in middle childhood.

The divergence in the moments of emergence for face-trait character and behavior judgments implies that, though face-trait character judgments are more directly evaluative and may therefore be simple for young children (perhaps even with innate abilities in infancy, Jessen & Grossmann, 2016), face-trait behavior judgments are more indirect and may first require a proficient understanding of traits and personality. That is, children may need to take the outputs from a face-trait character evaluation as their inputs for a face-trait behavior judgment. Such a mechanism is suggested by previous literature showing that children's behavior-to-behavior predictions also rely on trait reasoning (e.g., Chen et al., 2016). Indeed, the current results suggest that what is developing in children's capacity to make behavior judgments from face-traits is the developing concordance—both at the aggregate level and at the individual-level—between character and behavior judgments.

More generally, the results imply that children may need more experience, such as tracking associations between behaviors directed *toward* targets and the face-trait appearance of those targets, before making reliable explicit behavior judgments from faces. That is, although it is true that there is no reliable association between face-trait appearance and the behaviors *of* targets (Olivola et al., 2014), there is consistent evidence that perceivers moderate their behaviors *toward* targets based on a target's appearance (Todorov et al., 2015). As such, the association between, for example, a target's attractive facial appearance and positive behaviors toward that target (Langlois et al., 2000) or a target's competent face-trait appearance and positive support in elections (Todorov et al., 2005), may lead children to associate a certain face-trait appearance with a certain directed behavior.

Acquiring such an association between face-trait appearance and behavior directed toward a target may subsequently lead children in our studies to overgeneralize the association and offer judgments linking face-traits and behavior performed *by* a target. For example, children may come to notice that trustworthy-looking faces are more likely to receive help from others (and therefore judge the faces to be more likely to "help others" in turn),

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

dominant-looking faces are more likely to receive deference (and therefore judge them to be more likely to “decide which game to play”), and competent-looking faces are more likely to be praised for their abilities (and therefore judge them to be more likely to “sing songs well”). In other words, it is possible that one mechanism of developmental change comes from experience tracking associations between face-traits and behaviors directed toward targets, which then becomes overgeneralized to judgments about the behaviors of targets.

Of course, additional explanations of the observed developmental change are also possible. For instance, it may be that children observe extreme exemplars of face-trait appearances and behaviors in film or fiction and infer that an extremely dominant-looking face is always villainous and performs bad behaviors, whereas an extremely trustworthy-looking face is always heroic and performs good behaviors. To our knowledge, no studies have provided a direct test of this prediction for children’s perceptions of character and behaviors. Nevertheless, previous work shows a relationship between facial appearance and villainous behaviors in classic films (Croley, Reese, & Wagner, 2017), suggesting a plausible role for extreme fictional representations in shaping the emergence and persistence of children’s face-based judgments.

A second possibility is that children may overgeneralize emotion expressions, such that they perceive trustworthy, competent, or submissive faces to be happy and subsequently use this perception of “happy” as a more reliable cue to a target’s behaviors. Indeed, many of the positive face-traits are correlated with (although not redundant with) positive emotional expressions (Todorov et al., 2013; Todorov, Said, Engell, & Oosterhof, 2008). Children may therefore develop and maintain face-trait judgments, in part, because of the overlap between relatively uninformative face-traits and relatively more meaningful emotion cues.

Across these possible mechanisms, the continuous linear trend of increasing consistency in face-trait judgments suggests continuous maturation throughout childhood and even into adulthood. Thus, while research on adults’ face-trait behavior judgments (e.g., Todorov, 2017) may take for granted the speed and consistency of such judgments, the capacity for reliable face-trait behavior judgments may, in fact, be both difficult and gradually emerging. Further research examining the shift toward consistent behavior judgments occurring in early and middle childhood will be necessary to develop a more complete understanding of the social and cognitive inputs required for consequential face-trait judgments.

Remaining Questions and Limitations

The current experiments assessed judgments across multiple face-traits (trustworthiness, dominance, and competence) to newly show that children’s face-trait behavior judgments are not limited to competence (Antonakis & Dalgas, 2009), and their face-trait behavior performance is not limited to trust (Ewing et al., 2015). Rather, the consequences of children’s face-trait judgments appear far-reaching and robust across fundamental dimensions of social evaluations (Fiske et al., 2002; Oosterhof & Todorov, 2008). However, the current studies asked children to judge behaviors that were specifically matched to the face-trait being manipulated (e.g., to judge the relative likelihood of strength behaviors for dominant-looking faces). It is therefore unknown whether children at these young ages are sensitive to the *specificity* of the match between behaviors and face-trait information, or whether they make face-trait judgments based on more generalized affective preferences (e.g., halo effects, Brosseau-Liard & Birch, 2010; Cain et al., 1997). That is, would children judge that a trustworthy-looking face is also more likely to display competence (e.g., sing songs well) and dominance (e.g., pick up heavy boxes) because all of these behaviors are perceived to be relatively positive? Or are children’s face-based behavior judgments specific to the relevant face-trait?

Similarly, Experiments 3 and 4 demonstrated that children guide their own prosocial behaviors in response to face-traits, but the extent to which children are sensitive to the relevance of the specific face-trait when performing their behaviors is unknown. Would children prefer to learn from a competent-looking face, invest in a trustworthy-looking face, and be on a sports team with a dominant-looking face, but show no learning, investing, or sport team preferences toward faces with irrelevant features? Supplemental analyses of the current results suggest that children may not be particularly sensitive to the distinction between the appearance of dominance and trustworthiness when making generalized gift-giving decisions. However, it remains possible that children are, in fact, sensitive to the distinction across face-traits when performing more trait-specific behaviors (e.g., learning). Just as past research has explored the specificity of children’s behavior inferences from information about a target’s past behaviors (e.g., Hermes et al., 2015) or attractiveness (Langlois et al., 2000; Langlois & Stephan, 2006), it will be fruitful to understand the specificity of children’s behavior judgments and performance from relatively arbitrary and subtle face-trait information.

Additionally, the present research did not seek to contrast the relatively superficial “surface” cues of face-trait information with other “deeper” cues that are perceived to be more diagnostic of a target’s behavior and character. Research with adults has suggested that the influence of face-trait information in character and behavior judgments may persist even beyond relatively more diagnostic cues such as past behavior (Blair, Chapleau, & Judd, 2005; Suzuki, 2018), explicit information about character (Rule, Tskhay, Freeman, & Ambady, 2014), or social group membership (Kubota & Ito, 2007). In other words, face-trait information appears to have a privileged status in adults’ social judgments. However, given the current evidence that children use face-trait information in behavior judgments only after age 5, it is possible that such prioritization of face-trait cues does not reflect an innate human tendency, but rather may be one that is acquired gradually after tracking the relationships between facial appearance and behaviors directed toward others. Future research could examine age-related change in the relative roles of face-trait cues over competing information, including the relatively “deeper” cues of past behavior (e.g., Hermes et al., 2015) or social group membership (Charlesworth & Banaji, 2018; Langlois & Stephan, 2006; Rennels & Langlois, 2014). Such research would provide insights into the developmental boundary conditions on the prioritization of face-trait cues in shaping social judgments.

Finally, the current experiments used computer-generated faces that have been extensively validated to communicate latent face-traits (Todorov et al., 2008, 2013, 2015; Todorov & Oosterhof, 2011). The clarity and extremity of these stimuli is crucial to identify the specific facial features that correspond to the perception and judgments of character and behaviors. Yet these faces do not necessarily reflect the real-world complexity of children's and adult's exposure to, and judgments of, faces. Nevertheless, previous research (as well as the present Experiment 2) suggests robustness in children's and adult's judgments between computer-generated faces and more naturalistic stimuli, including real images and faces of other species (Cogsdill & Banaji, 2015). While the current findings are therefore unlikely to be an artifact of the stimuli, future examinations with naturalistic faces will be helpful in addressing whether naturalistic designs alter the age-related pattern in children's character and behavior judgments from face-traits.

Conclusion

The early emerging use of face-traits in social judgments extends beyond character evaluations. From at least age 5, face-trait information also guides children's judgments about how others will behave, as well as children's decisions about how to behave toward others. While research has shown that adults use face-traits in many important social decisions, the present work provides the first evidence of the extended developmental emergence of face-trait behavior judgments across multiple traits and behaviors. The findings suggest that, although later-emerging than character evaluations, the use of face-trait information in behavior judgments appears even before formal schooling, indicating an early understanding of the social implications of face-traits. More generally, the research highlights the benefit of a developmental approach in understanding patterns of change and inputs into perceivers' explicit social judgments and interactions.

References

Alvarez, J. M., Ruble, D. N., & Bolger, N. (2001). Trait understanding or evaluative reasoning? An analysis of children's behavioral predictions. *Child Development, 72,* 1409–1425. http://dx.doi.org/10.1111/1467-8624.00356

Antonakis, J., & Dalgas, O. (2009). Predicting elections: Child's play! *Science, 323,* 1183. http://dx.doi.org/10.1126/science.1167748

Baron, A. S., & Banaji, M. R. (2006). The development of implicit attitudes. Evidence of race evaluations from ages 6 and 10 and adulthood. *Psychological Science, 17,* 53–58. http://dx.doi.org/10.1111/j.1467-9280.2005.01664.x

Bascandziev, I., & Harris, P. L. (2016). The beautiful and the accurate: Are children's selective trust decisions biased? *Journal of Experimental Child Psychology, 152,* 92–105. http://dx.doi.org/10.1016/j.jecp.2016.06.017

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67,* 1–48. http://dx.doi.org/10.18637/jss.v067.i01

Blair, I. V., Chapleau, K. M., & Judd, C. M. (2005). The use of Afrocentric features as cues for judgment in the presence of diagnostic information. *European Journal of Social Psychology, 35,* 59–68. http://dx.doi.org/10.1002/ejsp.232

Blair, I. V., Judd, C. M., & Chapleau, K. M. (2004). The influence of Afrocentric facial features in criminal sentencing. *Psychological Science, 15,* 674–679. http://dx.doi.org/10.1111/j.0956-7976.2004.00739.x

Brosseau-Liard, P. E., & Birch, S. A. J. (2010). 'I bet you know more and are nicer too!': What children infer from others' accuracy. *Developmental Science, 13,* 772–778. http://dx.doi.org/10.1111/j.1467-7687.2009.00932.x

Cain, K. M., Heyman, G. D., & Walker, M. E. (1997). Preschoolers' ability to make dispositional predictions within and across domains. *Social Development, 6,* 53–75. http://dx.doi.org/10.1111/j.1467-9507.1997.tb00094.x

Charlesworth, T. E. S., & Banaji, M. R. (2018). *Face-trait and face-race cues in adults' and children's social evaluations*. Manuscript submitted for publication.

Chen, E. E., Corriveau, K. H., & Harris, P. L. (2016). Person perception in young children across two cultures. *Journal of Cognition and Development, 17,* 447–467. http://dx.doi.org/10.1080/15248372.2015.1068778

Cogsdill, E. J., & Banaji, M. R. (2015). Face-trait inferences show robust child-adult agreement: Evidence from three types of faces. *Journal of Experimental Social Psychology, 60,* 150–156. http://dx.doi.org/10.1016/j.jesp.2015.05.007

Cogsdill, E. J., Todorov, A. T., Spelke, E. S., & Banaji, M. R. (2014). Inferring character from faces: A developmental study. *Psychological Science, 25,* 1132–1139. http://dx.doi.org/10.1177/0956797614523297

Croley, J. A., Reese, V., & Wagner, R. F., Jr. (2017). Dermatologic features of classic movie villains: The face of evil. *Journal of the American Medical Association Dermatology, 153,* 559–564. http://dx.doi.org/10.1001/jamadermatol.2016.5979

Duarte, J., Siegel, S., & Young, L. (2012). Trust and credit: The role of appearance in peer-to-peer lending. *Review of Financial Studies, 25,* 2455–2484. http://dx.doi.org/10.1093/rfs/hhs071

Ewing, L., Caulfield, F., Read, A., & Rhodes, G. (2015). Perceived trustworthiness of faces drives trust behaviour in children. *Developmental Science, 18,* 327–334. http://dx.doi.org/10.1111/desc.12218

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. http://dx.doi.org/10.1080/17405620601046832

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82,* 878–902. http://dx.doi.org/10.1037/0022-3514.82.6.878

Freeman, J. B., & Johnson, K. L. (2016). More than meets the eye: Split-second social perception. *Trends in Cognitive Sciences, 20,* 362–374. http://dx.doi.org/10.1016/j.tics.2016.03.003

Fusaro, M., Corriveau, K. H., & Harris, P. L. (2011). The good, the strong, and the accurate: Preschoolers' evaluations of informant attributes. *Journal of Experimental Child Psychology, 110,* 561–574. http://dx.doi.org/10.1016/j.jecp.2011.06.008

Graham, J. R., Harvey, C. R., & Puria, M. (2017). A corporate beauty contest. *Management Science, 63,* 3044–3056. http://dx.doi.org/10.1287/mnsc.2016.2484

Green, P., & Macleod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution, 7,* 493–498. http://dx.doi.org/10.1111/2041-210X.12504

Hermes, J., Behne, T., & Rakoczy, H. (2015). The role of trait reasoning in young children's selective trust. *Developmental Psychology, 51,* 1574–1587. http://dx.doi.org/10.1037/dev0000042

Hermes, J., Behne, T., Studte, K., Zeyen, A. M., Gräfenhain, M., & Rakoczy, H. (2016). Selective cooperation in early childhood - How to choose models and partners. *PLoS ONE, 11,* e0160881. http://dx.doi.org/10.1371/journal.pone.0160881

Heyman, G. D. (2009). Children's reasoning about traits. *Advances in Child Development and Behavior, 37,* 105–143. http://dx.doi.org/10.1016/S0065-2407(09)03703-3

Jessen, S., & Grossmann, T. (2016). Neural and behavioral evidence for infants' sensitivity to the trustworthiness of faces. *Journal of Cognitive Neuroscience, 28,* 1728–1736. http://dx.doi.org/10.1162/jocn_a_00999

Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2012). "Native" objects and collaborators: Infants' object choices and acts of giving reflect favor for native over foreign speakers. *Journal of Cognition and Development, 13,* 67–81. http://dx.doi.org/10.1080/15248372.2011.567200

Kreft, I., & de Leeuw, J. (1998). *Introducing multilevel modeling*. London, UK: SAGE Publications. http://dx.doi.org/10.4135/9781849209366

Kubota, J. T., & Ito, T. A. (2007). Multiple cues in social perception: The time course of processing race and facial expression. *Journal of Experimental*

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Social Psychology, 43, 738–752. http://dx.doi.org/10.1016/j.jesp.2006.10.023

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. http://dx.doi.org/10.1037/0033-2909.126.3.390

Langlois, J. H., Roggman, L. A., Casey, R. J., Ritter, J. M., Rieser-Danner, L. A., & Jenkins, V. Y. (1987). Infant preferences for attractive faces: Rudiments of a stereotype? *Developmental Psychology, 23*, 363–369. http://dx.doi.org/10.1037/0012-1649.23.3.363

Langlois, J. H., & Stephan, C. (2006). The effects of physical attractiveness and ethnicity on children's behavioral attributions and peer preferences. *Child Development, 48*, 1694–1698. http://dx.doi.org/10.2307/1128538

Lawson, C., Lenz, G. S., Baker, A., & Myers, M. (2010). Looking like a winner: Candidate appearance and electoral success in new democracies. *World Politics, 62*, 561–593. http://dx.doi.org/10.1017/S0043887110000195

Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology, 51*, 93–120. http://dx.doi.org/10.1146/annurev.psych.51.1.93

McAuliffe, K., Raihani, N. J., & Dunham, Y. (2017). Children are sensitive to norms of giving. *Cognition, 167*, 151–159. http://dx.doi.org/10.1016/j.cognition.2017.01.006

Oh, D., Buck, E. A., & Todorov, A. (2019). Revealing hidden gender biases in competence impressions of faces. *Psychological Science, 30*, 65–79. http://dx.doi.org/10.1177/0956797618813092

Olivola, C. Y., Funk, F., & Todorov, A. (2014). Social attributions from faces bias human choices. *Trends in Cognitive Sciences, 18*, 566–570. http://dx.doi.org/10.1016/j.tics.2014.09.007

Olivola, C. Y., & Todorov, A. (2010). Fooled by first impressions? Reexamining the diagnostic value of appearance-based inferences. *Journal of Experimental Social Psychology, 46*, 315–324. http://dx.doi.org/10.1016/j.jesp.2009.12.002

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the United States of America, 105*, 11087–11092. http://dx.doi.org/10.1073/pnas.0805664105

Porter, S., ten Brinke, L., & Gustaw, C. (2010). Dangerous decisions: The impact of first impressions of trustworthiness on the evaluation of legal evidence and defendant culpability. *Psychology, Crime & Law, 16*, 477–491. http://dx.doi.org/10.1080/10683160902926141

Raabe, T., & Beelmann, A. (2011). Development of ethnic, racial, and national prejudice in childhood and adolescence: A multinational meta-analysis of age differences. *Child Development, 82*, 1715–1737. http://dx.doi.org/10.1111/j.1467-8624.2011.01668.x

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Rennels, J. L., & Langlois, J. H. (2014). Children's attractiveness, gender, and race biases: A comparison of their strength and generality. *Child Development, 85*, 1401–1418. http://dx.doi.org/10.1111/cdev.12226

Rezlescu, C., Duchaine, B., Olivola, C. Y., & Chater, N. (2012). Unfakeable facial configurations affect strategic choices in trust games with or without information about past behavior. *PLoS ONE, 7*, e34293. http://dx.doi.org/10.1371/journal.pone.0034293

Rholes, W. S., & Ruble, D. N. (1984). Children's understanding of dispositional characteristics of others. *Child Development, 55*, 550. http://dx.doi.org/10.2307/1129966

Rule, N. O., Tskhay, K. O., Freeman, J. B., & Ambady, N. (2014). On the interactive influence of facial appearance and explicit knowledge in social categorization. *European Journal of Social Psychology, 44*, 529–535. http://dx.doi.org/10.1002/ejsp.2043

Sobel, D. M., & Kushnir, T. (2013). Knowledge matters: How children evaluate the reliability of testimony as a process of rational inference. *Psychological Review, 120*, 779–797. http://dx.doi.org/10.1037/a0034191

Stipek, D. J., & Daniels, D. H. (1990). Children's use of dispositional attributions in predicting the performance and behavior of classmates. *Journal of Applied Developmental Psychology, 11*, 13–28. http://dx.doi.org/10.1016/0193-3973(90)90029-J

Suzuki, A. (2018). Persistent reliance on facial appearance among older adults when judging someone's trustworthiness. *The Journals of Gerontology Series B, Psychological Sciences and Social Sciences, 73*, 573–583.

Todorov, A. (2017). *Face value: The irresistible influence of firsts impressions*. Princeton, NJ: Princeton University Press.

Todorov, A., Dotsch, R., Porter, J. M., Oosterhof, N. N., & Falvello, V. B. (2013). Validation of data-driven computational models of social perception of faces. *Emotion, 13*, 724–738. http://dx.doi.org/10.1037/a0032335

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science, 308*, 1623–1626. http://dx.doi.org/10.1126/science.1110589

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology, 66*, 519–545. http://dx.doi.org/10.1146/annurev-psych-113011-143831

Todorov, A., & Oosterhof, N. (2011). Modeling social perception of faces. *IEEE Signal Processing Magazine, 28*, 117–122. http://dx.doi.org/10.1109/MSP.2010.940006

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition, 27*, 813–833. http://dx.doi.org/10.1521/soco.2009.27.6.813

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

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*, 1–48. http://dx.doi.org/10.18637/jss.v036.i03

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17*, 592–598. http://dx.doi.org/10.1111/j.1467-9280.2006.01750.x

Wilson, J. P., & Rule, N. O. (2015). Facial trustworthiness predicts extreme criminal-sentencing outcomes. *Psychological Science, 26*, 1325–1331. http://dx.doi.org/10.1177/0956797615590992

Received July 10, 2018
Revision received February 26, 2019
Accepted March 6, 2019 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.