The Development of Character Judgments From Faces

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th>Citation</th>
<th>Cogsdill, Emily. 2015. The Development of Character Judgments From Faces. Doctoral dissertation, Harvard University, Graduate School of Arts &amp; Sciences.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:17463958">http://nrs.harvard.edu/urn-3:HUL.InstRepos:17463958</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
The Development of Character Judgments from Faces

A dissertation presented

by

Emily Cogsdill

to

the Department of Psychology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Psychology

Harvard University

Cambridge, Massachusetts

February 2015
The Development of Character Judgments from Faces

ABSTRACT

First impressions play a central role in human social interaction. In particular, the face is a rich source of information that perceivers use in making both initial and lasting character judgments. Despite the large and growing body of work demonstrating that these judgments affect outcomes in domains as crucial as elections and criminal sentencing, little remains known about the ontogenetic origins of this consequential aspect of human social cognition. The purpose of this dissertation is to provide a set of early investigations into the development of face-to-trait inferences. Paper 1 demonstrates that, like adults, even children as young as 3-4 years of age provide consistent trait judgments when asked to judge two-dimensional computer-generated face images, suggesting that this general ability is so fundamental as to emerge strongly at the earliest ages tested. Paper 2 shows that this propensity is so deeply ingrained at an early age that similar consensus across the lifespan emerges in response to static faces belonging to adults, children, and even rhesus macaques. Paper 3 investigates the potential consequences of these judgments, showing that face-based character assessments influence attributions of trait-relevant behaviors and even cause children to modulate their own behaviors towards others. The findings of this dissertation clearly illustrate that face-to-trait inference emerges early in development, and is therefore a fundamental element of human social cognition with important consequences throughout the lifespan.
# TABLE OF CONTENTS

ACKNOWLEDGMENTS ................................................................................................................. v  

INTRODUCTION ........................................................................................................................... 1  
Faces as a cue to personality ....................................................................................................... 5  
Development ............................................................................................................................... 7  
Overview of dissertation ............................................................................................................ 12  
A note on accuracy ..................................................................................................................... 14  

PAPER I: Inferring character from faces: A developmental study ........................................ 17  
Abstract .................................................................................................................................... 17  
Experiment 1 ............................................................................................................................... 21  
Experiment 2 ............................................................................................................................... 25  
General Discussion .................................................................................................................... 29  

PAPER II: Face-trait inferences show robust child-adult agreement: Evidence from three  

types of faces ............................................................................................................................... 31  
Abstract .................................................................................................................................... 31  
Experiment 1 ............................................................................................................................... 35  
Experiment 2 ............................................................................................................................... 47  
General Discussion .................................................................................................................... 54  

PAPER III: Children modulate their behavioral attributions and manifest behaviors based  
on the facial appearance of others ............................................................................................. 58  
Abstract .................................................................................................................................... 58  
Experiment 1: From Trait Inferences to Behavior Attributions ............................................. 62  
Experiments 2-5: Face-Based Modulation of Prosocial Behaviors ....................................... 81  
Experiment 2 ............................................................................................................................... 84  
Experiment 3 ............................................................................................................................... 88  
Experiment 4 ............................................................................................................................... 92  
Experiment 5 ............................................................................................................................... 95  

CONCLUSION .............................................................................................................................. 104  
REFERENCES ............................................................................................................................. 114
ACKNOWLEDGMENTS

For advice and assistance in preparing this dissertation, I thank the members of my committee: Mahzarin Banaji, Elizabeth Spelke, Ken Nakayama, and Max Krasnow.

For their financial support of this work, I thank Harvard University, the American Psychological Foundation, the E. J. Safra Center for the Study of Ethics, the Santa Fe Institute, and the National Institutes of Health.

For their assistance discussing ideas, reviewing manuscripts, and enduring hours of watching me talk and gesticulate wildly at pictures of faces and bar graphs, I thank my colleagues in the Implicit Social Cognition Laboratory and the Laboratory for Developmental Studies.

For their elbow grease and relentlessly positive attitudes, I thank the research assistants with whom I have been fortunate to work these past three years: Amelia Clark, Sarah Coughlon, Ann Finkel, Judy Jiao, Neela Kaushik, Bryan LeBlanc, Carol Meuth, and Elyse Neubauer.

For their shared mischief and endless moral support, I thank my dear friends.

For their unconditional love and encouragement, I thank my family.

Finally, for her wisdom, grit, and unfailing mentorship, I thank Mahzarin Banaji.
INTRODUCTION
Humans are profoundly social creatures. First impressions play a critical role in social interaction, and perceivers readily use appearance cues to inform judgments of personality traits (e.g., Hassin & Trope, 2000; Zebrowitz, 1997; Zebrowitz & Montepare, 2005). Most fundamentally, such judgments inform our inclinations to approach or avoid others (Todorov, 2008), and from an evolutionary standpoint it stands to reason that humans would have developed perceptual systems to aid in survival by determining helpful collaborators or dangerous threats in the environment. Whatever their phylogenetic origin, these appearance-based first impressions persist in the modern world. Advances in psychological research methods have further demonstrated that character judgments in response to facial appearances emerge spontaneously (Todorov & Uleman, 2002, 2003) and rapidly (e.g., Bar et al. 2006), with even as little as 33 milliseconds of exposure to masked images being sufficient to elicit trait judgments (Todorov et al., 2009). Moreover, these judgments remain relatively unchanged even with unlimited exposure to faces (Ballew & Todorov, 2007). First impressions are thus at the heart of the perpetual challenge of understanding one another (Gilbert, 1998), an endeavor that is essential to cooperating and functioning in modern society.

It is therefore appropriate that decades of psychological research, and much of the field of social psychology in general, has focused on studying and articulating the processes that guide this quest for mutual understanding. Modern research makes extensive use of sophisticated techniques like neuroimaging and data-driven statistical models to understand how appearances create our perceptions of one another. The origins of such work can be traced back to such influential researchers as Asch (1946) and Heider (1958), whose experiments and theories in the mid-20th century fueled a broad
interest in understanding how humans generate causal attributions to understand each others’ behavior in everyday life – i.e., how they form impressions of one another. This conception of social reasoning as an extension of basic causal reasoning inspired a long tradition of research on person perception. Attribution theory and the work it generated centered on the question of how behaviors lead to attributions of internal (i.e., dispositional) versus external (situational) factors. The basic process of attribution is simple enough: through their behaviors and appearance, actors provide some number of cues that perceivers use to infer both the invariant, stable characteristics of the actor, as well as the variable circumstances of the surrounding environment. But what is the nature of cues that actors provide, and which are the ones that perceivers actually use? To what degree do individual perceivers use these cues consistently versus idiosyncratically? And are any of these impressions even accurate? The eternal task of research in social cognition is to gradually fill in the details surrounding these central mysteries of the human social mind.

Much of this work has focused on how perceivers use behavioral information to extract information about dispositions. However, from the days of Darwin and even Aristotle, scientists have been enamored with the idea of finding hidden clues to character contained within our physical appearance. While in modern times the quest to unlock these secrets of the human spirit are carried out through validated tests of intelligence and personality and biotechnological tools like genome sequencing, for centuries people relied on outward appearance, such as facial structures, to attempt to generate such insights. Aristotle is said to have written about the connection between facial features and personality (see, e.g., Zebrowitz, 1997), foreshadowing a practice that came to be known
centuries later as *physiognomy* (Alley, 1988). The belief that physical appearance contains clues to internal character traits gained further traction in the 17th century through the work of Johann Caspar Lavater, a Swiss physician whose *Essays on Physiognomy* captured the mind of the public (see Shookman, 1993). Physiognomy, as well as its pseudoscientific cousin, phrenology, saw renewed interest in the earliest 20th century, during which time scientists like Cesare Lombroso (1895) took a particular interest in trying to determine criminality based on facial features present even at birth. Even in the present day, despite myriad advances in science and education worldwide, the idea that faces hold the secret to underlying behavioral tendencies persists in the minds of many. One survey of Israelis in the year 2000 revealed that a full two-thirds of respondents considered it possible to know at least “a few traits” from somebody’s facial appearance (Hassin & Trope, 2000). The allure of faces and the pseudoscience of decoding hidden meanings from its underlying structures thus continues to influence even the explicit judgments people make, to say nothing of the multitude of implicit everyday judgments that elude conscious awareness.

Following the failure of researchers to find meaningful and robust connections between appearance and character traits in the early 20th century (though much work continues to explore the predictive validity of appearance; see “A note on accuracy” on page 13), starting in around the 1950s psychologists studying person perception began to switch their focus from how accurate certain features are in predicting personality to simply focusing on what cues people use to render these judgments – in the words of Gilbert (1998), a shift “toward questions of *how* and away from questions of *how well*” (see also Bruner & Tagiuri, 1954). Following in this historic trend, a great deal of
experimental work on first impressions made use of “zero acquaintance” paradigms (e.g., Borkenau & Liebler, 1992), in which perceivers are typically asked to render judgments to unfamiliar others based on some sample of exposure to them, often in the form of a photograph or a short video. A particularly influential source of such work is that which has studied how perceivers use “thin slices” – that is, very short exposure to non-verbal appearance or behaviors, or even incidental scenes such as pictures of a person’s bedroom – to render important and lasting judgments about other people (for a review, see Ambady & Weisbuch, 2010). Decades of psychological research have used such methods to attempt to gain purchase on the central questions surrounding the nature of first impressions, as well as their ultimate consequences for social interaction.

**Faces as a cue to personality**

While an abundance of research has investigated the inferences that people make based on stimuli such as short videos of unfamiliar others (see Ambady & Rosenthal, 1992 for a meta-analysis), the work contained in this dissertation focuses on inferences that perceivers make from faces. It is well established that faces are rich sources of information, and humans are adept at rapidly extracting information from the faces of others. Within milliseconds, perceivers are able to categorize faces into social groups such as sex (Valdés-Conroy et al., 2014) and race (Cunningham et al., Ito, Thompson, & Cacioppo, 2004), and specialized neural systems encode group membership in these domains (Contreras, Banaji, & Mitchell, 2013). Perceivers also readily attribute personality traits to others based on mere facial structure, and they do so rapidly (Bar, Neta, & Linz, 2006; Willis & Todorov, 2006) and with broad consensus (e.g., Rule,
Krendl, Ivcevic, & Ambady, 2013), with similar judgments persisting even across different cultures (Rule & Ambady, 2010). Moreover, in a fashion similar to more basic processes like face recognition and categorization, even these face-based character judgments have been shown to be neurally encoded, with the amygdala showing responses that correspond with consensus judgments of facial trustworthiness (Todorov & Engell, 2008).

Research studying face-based character judgments is given particular urgency in light of work showing their numerous real-world consequences. One of the most well-documented effects is that of facial neoteny (i.e., resemblance to children), commonly known as “babyfacedness,” which has been shown to elicit judgments of warmth, honesty, and incompetence (Berry & Zebrowitz-McArthur, 1985). The influence of babyfacedness has even been shown to affect sentencing outcomes in small claims courts (Zebrowitz & McDonald, 1991). Other work in the past decade has made the provocative discovery that United States Congressional elections have outcomes that can be predicted by facial competence of candidates with a degree of accuracy that many find alarming in the context of a Democratic society (Ballew & Todorov, 2007; Todorov, Mandisodza, Goren, & Hall, 2005). Structural resemblance to Afrocentric features predicts criminal sentencing (Blair, Judd, & Chapleau, 2004), and facial cues signaling dominance predict career success in the military (Mueller & Mazur, 1996), law (Rule & Ambady, 2011) and in business (Wong, Ormiston, & Haselhuhn, 2011). These and other such examples demonstrate clearly that inferences from faces have powerful consequences that pervade society. The downstream effects of first impressions are particularly impressive given the social mores that implore us not to “judge a book by its cover,” to cite an adage that is
both ubiquitous and routinely disregarded. Such work revealing the consequential nature of face-to-trait inferences highlights the need for psychologists to continue to work towards understanding their underlying social cognitive mechanisms.

**Development**

One way of studying the mechanisms that drive face-to-trait judgments is to study their ontogeny, specifically by asking the question of when such judgments can first be observed in development. Research in developmental psychology has revealed that the basic elements of human cognition can be observed in infants, whose core cognitive capacities enable them to perceive and reason about objects (e.g., Baillargeon, 2004), interpret actions (e.g., Woodward, 1998), understand number concepts (e.g., Feigenson, Dehaene, & Spelke, 2004), and reason about geometric forms (e.g., Spelke, Lee, & Izard, 2010). Such work shows clearly that the basic building blocks of adults’ ability to perform complex cognitive feats, such as arithmetic or navigation, have their earliest roots in infancy.

In a similar vein, the burgeoning field of social cognitive development continues to reveal the primacy of many different capacities in the social domain as well. Cognitive abilities that are fundamental to understanding others, such as theory of mind, have been shown to develop in children as young as three years of age (Wellman, Cross, & Watson, 2001), and some have even argued that infants possess the rudiments of this fundamental social skill in the form of understanding false beliefs (Onishi & Baillargeon, 2005). Early childhood and even infancy sees the emergence of many other basic abilities, such as goal understanding (Woodward, 1998, 1999), gaze following and social referencing
(Baldwin, 1991; Johnson, Slaughter, & Carey, 2000; Moses, Baldwin, Rosicky, & Tidball, 2001), joint intentionality (Grafenhein, Behne, Carpenter, & Tomasello, 2009; Tomasello & Carpenter, 2007), and imitation (Gergely, Bekkering, & Király, 2002; Király, Csibra, & Gergely, 2013). Even complex forms of social cognition have been observed early in development. For example, young children demonstrate implicit bias based on race (Dunham, Baron, & Banaji, 2008; Dunham, Chen, & Banaji, 2013) and gender (Cvencek, Greenwald, & Meltzoff, 2011) and show signs of early-developing biases based on religious beliefs (Heiphetz, Spelke, & Banaji, 2013) and even pure chance (Olson, Dunham, Dweck, Spelke, & Banaji, 2008).

Despite this growing body of research, very little is known about how children attribute traits to faces. Two papers published in the 1980s investigated whether children share the same babyface stereotypes that adults do (Keating & Bai, 1986; Montepare & Zebrowitz-McArthur, 1989); however, these studies were highly limited both in their sample sizes and the ages of their participants, who ranged from only 3 to 7 years of age and did not include older children or adults. This early work was also limited by the measures they used to probe trait judgments, many of which suffered from conflating judgments of dominance and generalized valence. Little if any other work addressed the question of how children derive character inferences from faces until 2009, when a one-page report published in Science made the intriguing observation that children’s selections of whom to be the captain of a fictional ship, when presented with options consisting of faces of electoral candidates, tracked closely with actual results of those elections (Antonakis & Dalgas, 2009). Even more recently, in 2014 researchers made a more substantive investigation into how children use facial appearance to decide how to
behave towards their partners in an economic Trust Game (Ewing, Caulfield, Read, & Rhodes, 2014), providing an early investigation into the question of behavioral modulation that is discussed more thoroughly in the third paper of this dissertation (page 54). However, these papers represent nearly, if not all, of the extant research studying children’s character judgments of others based on facial appearance. This striking lacuna in social cognitive development research suggests a clear need for more thorough and rigorous investigations using controlled experiments and large samples of both children and adults.

Although the existence of prior work studying children’s face-to-trait inferences is scant, a wealth of research studying face perception and trait understanding early in development can provide some clues as to which social cognitive faculties young children might possess. To begin with, a great deal is known about the primacy of facial perception skills, which in many respects reach fully adult-like proficiency within the first year of life. Infants in their first year of life can discriminate faces from nonfaces, and even prefer to look at stimuli with face-like configurations of features (Cassia, Turati, & Simion, 2004; Farroni et al., 2006; Johnson, Dziurawiec, Ellis, & Morton, 1991; Simion et al., 2002). They are also able to recognize facial identities (Moulson et al., 2011; Pascalis, Haan, & Nelson, 1998; though see also Germaine, Duchaine, & Nakayama, 2011, whose findings suggest that facial recognition may not reach its full potential until well into adulthood), follow other faces’ gazes (Corkum & Moore, 1998; Scaife & Bruner, 1975), and even interpret social and non-social actions using gaze direction (Hood, Willen, & Driver, 1998; Johnson, Slaughter, & Carey, 1998; Farroni, Massaccesi, Pividori & Johnson, 2004).
Infants also show a robust tendency to prefer certain types of individuals based on their facial appearances. Within the first year of life, babies prefer to look at faces based on differences in attractiveness (Rubenstein, Kalakanis, & Langlois, 1999; Langlois et al., 1987; Langlois, Ritter, Roggman, & Vaughn, 1991; Ramsey et al., 2004), gender (Quinn, Yahr, Kuhn, Slater, & Pascalis, 2002), and ethnicity (Kelly et al., 2005; Bar-Haim et al., 2006). Finally, although identifying expressions of facial emotion is more difficult than simply perceiving faces, this ability begins to emerge by two years of age (Widen & Russell, 2003), if not earlier (Walker-Andrews, 1997). Taken together, these findings suggest that the face perception skills necessary for face-trait inferences are certainly in place before preschool, and possibly even earlier.

The crux of children’s emerging abilities to make facial trait inferences may therefore lie in their abilities to attribute personality traits to others. The existing research on this topic is mixed. Children younger than seven years of age are less likely than older children to use trait words to describe themselves or other people (e.g., Eder, 1989), and they tend to fail simple behavior-to-behavior prediction tasks (e.g., Berndt & Heller, 1985; Ferguson, van Roozendaal, & Rule, 1986; Rholes & Ruble, 1984). However, there may also be reason to believe that young children possess some understanding of personality traits. Children as young as 3-4 years of age have been shown to reason about other people’s mental states, using knowledge about others’ beliefs and desires to predict their future behavior (Wellman, 1992; see also Wellman, Cross, & Watson, 2001 for a meta-analysis). Additionally, young children use categorical (but not dimensional) trait information more readily than physical appearance when attributing preferences to fictional characters (Heyman & Gelman, 1999, 2000; Gonzalez, Zosuls, & Ruble, 2010).
Consequently, researchers have begun to suggest that, although they struggle to make behavior-to-behavior predictions, children as young as four years of age successfully make behavior-to-trait inferences and trait-to-behavior predictions (Liu, Gelman, & Wellman, 2007), although they may require greater amounts of information about the frequency of past behaviors in order to do so (Boseovski & Lee, 2006).

Other recent work showing that children use behaviors to infer competence in others (e.g., Harris, 2012) suggests that the failure to generate behavior-to-behavior predictions observed in prior research suggests not that children are unable to make such inferences, but that these inferences are less likely to be detected using verbal measures. In addition, further research has shown that even infants appear to use behavioral cues to infer enduring characteristics like prosociality (Hamlin, Wynn, & Bloom, 2007) and even dominance (Thompson, Frankenhuis, Ingold-Smith, & Carey, 2011; Mascaro & Csibra, 2012). Such work suggests that prior research that relied on verbal measures may have underestimated young children’s understanding of stable dispositions as a source of behavioral consistency across different situations.

Thus, as a result of their reliance on verbal measures, earlier studies may have underestimated young children’s understanding of personality traits. While a behavior might give us some indication of a person’s general tendencies (i.e., personality), however, faces – particularly static ones – should be a relatively poor source of such information. Related to this, Gonzalez et al. (2010) found that while the youngest children tested (under 6 years) readily used bodily similarities between characters to draw inferences about them, facial similarities alone were insufficient to drive these inferences.
This finding lent further support to the idea that face-based inferences may not be as intuitive to young children as they are to older children and adults.

For all of the above reasons, at the outset of this dissertation, we considered it possible that face-to-trait inferences should present a greater challenge, and therefore emerge later in development, than behavior-to-trait inferences. The alternative possibility, which the papers in this dissertation overwhelmingly support, is that face-to-trait inference is a social cognitive faculty that is so deeply ingrained in the mind of the growing child that adult-like judgments linking facial features to personality traits can be observed at the earliest ages we are currently able to test, at around three years of age.

**Overview of dissertation**

The overarching objective that ties together the studies reported here is to provide a robust and cohesive set of initial explorations into the development of face-to-trait inference and its possible consequences for children’s behavior. The first paper (Cogsdill, Todorov, Spelke, & Banaji, 2014) shows that, in stark contrast to any expectations that children might be reluctant to offere face-to-trait inferences, children readily attribute personality traits to faces in adult-like patterns. Furthermore, this work shows that they do so with even greater consistency when the judgment involves a basic “nice/mean” evaluation.

The second paper (Cogsdill & Banaji, 2014) extends these initial findings by asking children to apply basic valence judgments to faces that were selected to be more subtle in the traits that they signal to adult participants. Here, children viewed both a set of cropped and grayscaled human faces that were designed to appear relatively
homogenous as well as a set of rhesus macaque faces, which were expected to create a more difficult task given the species-specific nature of early-developing expertise in face perception.

The third and final paper (Cogsdill, Spelke, & Banaji, in preparation) probes the question of whether these trait inferences lead to differences in behavioral attributions as well as the manifest behaviors that children produce towards others based on facial appearance alone. The series of five experiments reported here show that children not only render “nice/mean” judgments to faces, thereby self-replicating earlier findings from both papers, but they also readily attribute trait-relevant behaviors to faces. For example, even the youngest children tested overwhelmingly identified trustworthy faces as being the ones who “help other people when they are in trouble,” or the competent-looking faces as “[knowing] how to sing a lot of different songs.” In the experiments that follow, children are asked to either select novel objects for themselves or decide which characters to give small “gifts,” such as an image of a laminated cookie. These methods yield the strikingly consistent finding that children readily modulate their own behaviors towards fictional characters based on facial appearances alone.

Taken together, the set of findings these papers report clearly demonstrate that children possess an adult-like propensity to render character judgments based on facial appearance alone. Moreover, this work shows that this ability is highly robust, emerging in response to static face images belonging to adults, children, and even rhesus macaque monkeys. Such judgments seemingly reveal the roots of adult social judgments, yet even in childhood they influence related judgments, such as behavioral attributions, and can even influence how children choose to behave towards others. The work contained within
this dissertation thus provides a collection of robust and consistent findings that lay a foundation for future work to more thoroughly characterize the scope of this important social cognitive faculty across the lifespan.

A note on accuracy

The topic of accuracy is one of such breadth, depth, and complexity that it has been deliberately omitted from the studies that comprise this dissertation, which focuses on the ontogenetic origins and developmental course of face-based character judgments early in life. However, the question of whether and to what degree social judgments based on facial characteristics are accurate has important implications for the phylogenetic origins of such stereotyped judgments. Certain aspects of physiognomy, such as inferences based on facial cues driven by testosterone (a hormone that can be linked to survival-relevant behaviors such as aggression), may have had fitness value in the human environment of evolutionary adaptedness, which in turn would contribute to selective pressures favoring a social perceptual system that makes early and consistent use of such judgments. Ultimately, a complete answer to the question of why modern humans have such rich and pervasive stereotypes linking many different types of features to personality traits will thus require some understanding of the accuracy (or lack thereof) of social perceptual mechanisms. Due to its broader importance in the field of impression formation and person perception research, the question of accuracy merits a brief overview here.

Much of the modern work on accuracy draws original inspiration from Brunswik’s lens model (Brunswik, 1956), an influential theory that described the processes underlying accuracy in personality judgments. Under this model, perceiver
accuracy is a function of cue validity (the degree to which features truly correspond with traits) and cue utilization (the degree to which perceivers use features to guide their judgments). This original theory has been extended to include distinctions between implicit and explicit self-perceptions and judgments of character (as in the dual-lens model; Hirschmüller, Egloff, Nestler, & Back, 2013). This general approach has inspired scores of papers in which participants are shown to formulate seemingly accurate judgments based on short videos (e.g., Borkenau & Liebler, 1992), or even uniquely modern analogues for social interactions such as online profiles (Back et al., 2010) and even email addresses (Back, Schmukle, & Egloff, 2008).

Such work has also focused on accuracy of inferences derived from faces alone, arguing that perceivers can use facial appearance to detect highly personal mental states such as sexual orientation (Rule, Ambady, Adams, & Macrae, 2008) and even religious beliefs (Rule, Garrett, & Ambady, 2010). A recent meta-analysis reported accuracy of 64.5% when categorizing a number of such perceptually ambiguous groups across 47 articles (Tskhay & Rule, 2013). Moreover, recent work has suggested a valid connection between cues to facial trustworthiness and trustworthy behavior in economics games (Stirrat & Perrett, 2010), as well as between dominance cues like facial width-to-height ratios and aggressive behaviors (e.g., Carré & McCormick, 2008). Significant associations between facial structure and measured personality have even been reported even for capuchin monkeys (Wilson et al., 2014). Such work suggests the possibility that some features may serve as valid cues to certain types of behavior. It stands to reason that such a connection, or a “kernel of truth” in such judgments, may provide a mechanism for the emergence of appearance-based biases to have emerged across evolutionary
history. However, despite the possible importance that such validity may have had in the phylogenetic origins of such judgments, the focus of the present investigation is instead on the ontogenetic origins of face-based character evaluations – i.e., its emergence and development across the lifespan. This approach necessitates that questions of accuracy be temporarily set aside, with a focus instead on measuring consensus of face-to-trait judgments from childhood through to adulthood.

Abstract

Human adults attribute traits to faces readily and with high consensus. In two experiments investigating the development of face-trait inference, adults and children ages 3-10 attributed trustworthiness, dominance and competence to pairs of faces. In Experiment 1, children as young as 3-4 years made face-trait attributions converging with those of adults and 5-6 year olds were at adult levels of consistency. In Experiment 2, children aged 3 and above consistently attributed the basic “mean/nice” evaluation to faces varying not only in trustworthiness but in dominance and competence. This research suggests that the predisposition to judge others using scant facial information that appears in adult-like forms early in childhood does not require prolonged social experience.
Faces command our attention and interest, and facial appearance has profound effects on social judgments (Todorov, Mende-Siedlecki, & Dotsch, 2013; Zebrowitz & Montepare, 2008). The speed and confidence with which we dispatch character assessments such as “trustworthy” or “competent” in response to a face is impressive. Face-to-trait inferences appear to be intuitive and automatic among human adults, and its development in early childhood is the focus here.

Prior research shows that face-trait inferences occur extremely rapidly, emerging within 50 milliseconds after exposure (Todorov, Pakrashi, & Oosterhof, 2009). Second, these character attributions show broad and cross-cultural consensus (Rule et al., 2010). Third, these judgments often result from overgeneralizing perceptions of facial configurations that signal ecologically valid information, such as emotional states (Said, Sebe & Todorov, 2009) and fitness (Zebrowitz & Rhodes, 2004). Finally, face-trait inferences occur even in consequential settings including criminal sentencing (Blair, Judd, & Chapleau, 2004), financial success (Rule & Ambady, 2008), and election outcomes (Todorov, Mandisodza, Goren, & Hall, 2005).

The present studies begin an exploration testing whether young children infer character traits like trust, competence, and dominance simply by looking at 2-D static images of faces, and if so how early in development they do so in an adult-like manner. If agreement in face-trait inferences emerges gradually across development, we might infer that they require prolonged social experience to manifest. If instead young children respond like adults, we would learn that face-to-trait character inferences are a

---

1 In this paper we do not address the veridicality of face-trait inference, as others have studied (e.g., Carré et al., 2009). Although this is an important topic, we focus centrally on the development of such inferences from the earliest ages that can be tested.
fundamental social cognitive capacity that emerges early in life. Thus, our investigation is simply one of whether children and adults make similar trait inferences based on the same faces.

We know that infants prefer to look at faces over non-faces and form preferences based on attractiveness, gender, and race (Bar-Haim et al., 2006; Langlois, 2000; Quinn, Yahr, Kuhn, Slater, & Pascalis, 2002; Ramsey et al., 2004; Simion et al., 2001). However, little is known about how older children use faces to make inferences about others’ character, and the existing research on this topic is mixed. Even though 3-4 year olds predict behavior from information about mental states (Wellman, Cross, & Watson, 2001), children under age seven usually fail simple behavior-to-behavior prediction tasks (Rholes & Ruble, 1984) and are less likely than older children to use trait words to describe people (Barenboim, 1981).

While some aspects of face-to-trait inference in children have been studied (see Antonakis & Dalgas, 2009; Clement et al., 2012; Keating & Bai, 1986; Montepare & Zebrowitz-McArthur, 1989) the present research explores the development of face-to-trait inferences within a wide age group and in a variety of domains. Importantly, our method enables us to test and compare responses between not only adults, but also 3-10 year old children. We explore face-based attributions of basic evaluations, like “nice/mean,” as well as assessments of more specific traits, like “strong” and “smart.”

**General Method**

Participants viewed computer-generated faces selected to be high or low on perceived *trustworthiness, dominance, or competence*. These extensively validated (Todorov et al.,
2013) faces were created in FaceGen based on data-driven, computational models of the respective traits (Oosterhof & Todorov, 2008; Todorov & Oosterhof, 2011;) In both experiments, we used three sets of faces, each of which included six distinct face identities. The three sets each contained faces that appeared high or low on a single trait (±3 SD in trustworthiness, dominance, and competence; see Figure 1.1).

![Figure 1.1. Sample stimuli from Experiment 1.](image)

In each trial, participants viewed two faces side-by-side, one high and one low in a trait. Face pairs appeared in three blocks (order counterbalanced across participants), each containing 9 trials in which all face pairs (low vs. high on trait) varying in all three traits appeared in a random order.

For each trial pair, participants identified the face that possessed a particular trait by answering short questions, e.g., “Which of these people is very nice/strong/smart?”
Children answered by pointing to one of the faces on the screen, and adults responded online by selecting a face. Faces and prompts were randomized with the constraint that anticipated responses appeared on either side of the screen with approximately equal frequency.

Experiment 1

Method

141 children (\(M_{\text{age}} = 6.5\), range = 3.1-10.11, 68 females) participated at local museums and in the laboratory\(^2\); 99 adults (\(M_{\text{age}} = 30.23\), range = 18-67, 54 females, 1 unspecified) participated online through SocialSci\(^3\). Participants attributed trustworthiness ("mean/nice"), dominance ("strong/not strong"), and competence ("smart/not smart") to pairs of faces. Each pair contained faces from the same set (trustworthiness, dominance or competence), with one face appearing high and the other low in that trait.

Results

Figures 1.2–1.4 summarize results from all ages and traits, with greater proportions of expected responses – i.e., those predicted based on prior data collected with these faces (e.g., trustworthy faces = “nice” and untrustworthy faces = "mean") – indicating stronger consensus. All age groups showed significant consensus compared to chance (50%) when attributing “mean/nice” (93%; Figure 1.2), “strong/not strong” (85%; Figure 1.3), and “smart/not smart” (76%; Figure 1.4).

Critically, all age groups made all three attributions with significant consensus, \(p < .001\), \(d > 1.08\). Although 3-4 year olds responded with robust and adult-like consensus

\(^2\) Ages for groups of child participants are notated as “years;months.”

\(^3\) www.socialsci.com.
(72% across all traits), they were less consistent than 5-6 year olds (81%), 7-10 year olds (88%), and adults (89%). These differences constituted a significant main effect of Age, $F(3,236) = 17.91, p < .001$. One-way ANOVAs followed by post-hoc tests using Sidak corrections for multiple comparisons were used to analyze age differences for each trait. This analysis revealed a pattern for both trustworthiness and dominance whereby 3-4 year olds were less consistent than all other age groups (all $ps < .01$, $ds > .59$), which were in turn equivalent to each other (all $ps > .23$, $ds < .40$).

Competence showed an altered developmental pattern whereby consensus primarily increased between the ages of 5-6 and 7-10. 3-4 year olds (68%) were identical to 5-6 year olds (66%; $p = 1.00$, $d = .07$) but less consistent than 7-10 year olds (84%, $p < .05$, $d = .64$) and marginally less consistent than adults (80%; $p = .08$, $d = .47$). In similar fashion, 5-6 year olds were less consistent than both 7-10 year olds ($p < .01$, $d = .67$) and adults ($p < .05$, $d = .51$). 7-10 year olds and adults attributed competence with similar consensus ($p = .91$, $d = .18$).

---

4 This result was statistically significant before correcting for multiple comparisons ($t(134) = 2.607, p = .01$).
Figure 1.2. Average rates of attributing “nice” to trustworthy and “mean” to untrustworthy faces. Error bars represent SEM.

Figure 1.3. Average rates of attributing “strong” to dominant and “not very strong” to submissive faces.
Figure 1.4. Average rates of attributing “smart” to competent and “not very smart” to incompetent faces.

Data were further analyzed using a 3 (trait: trustworthiness vs. dominance vs. competence) × 4 (age group: 3-4 year olds vs. 5-6 year olds vs. 7-10 year olds vs. adults) mixed model ANOVA with repeated measures on the first factor. This analysis revealed main effects of both Trait, $F(2.028^{5},472) = 42.66$, and Age Group, $F(3,236) = 18.09$, $p_s < .001$. These main effects were qualified by a Trait*Age Group interaction, $F(6,472) = 4.031$, $p < .01$. Within-subjects contrasts revealed that response accuracy was highest for judgments of trustworthiness (91.6% accuracy). This overall accuracy was significantly higher than that of dominance (81.5%; $F(1,236) = 54.24$, $p < .001$), which was higher than competence (74.4%; $F(1,236) = 10.10$, $p < .01$).

Overall, the data suggest that children’s face-trait inferences reach adult-like consensus at an impressively early age. For all three traits tested, children in the youngest

---

5 Degrees of freedom were adjusted using a Greenhouse-Geisser correction after Trait failed a test of sphericity, Mauchly’s $W = .726$, $p < .01$. 
age group responded with striking consistency greatly exceeding chance responding, although they were typically less consistent than older participants.

Although consensus was consistently high across all age groups and traits, the consensus that emerged in trustworthiness trials was significantly greater than that obtained in dominance and competence trials, suggesting that judgments of “mean” or “nice” might emerge uniquely early compared with other judgments. If true, such judgments might be fundamental to face-trait inference, and therefore broadly applied to faces varying in trait dimensions other than trustworthiness.

**Experiment 2**

Experiment 2 explored this possibility by testing whether “mean/nice” judgments emerge when viewing faces that vary in dominance and competence instead of trustworthiness. Given the primacy of valence evaluations in social judgments, children might robustly apply basic “nice/mean” judgments to faces varying in traits other than trustworthiness. If such evaluations rely on specific features varied in trustworthy/untrustworthy faces, however, consensus should not emerge when applying this global evaluation to other faces.

**Method**

A total of 203 children (Mage = 5;11, range = 3;1 - 10;8, 110 female, 2 of unspecified gender) participated at museums and in the laboratory, and 301 adults (Mage = 28.9, range = 18-72, 142 females, 6 unspecified) participated online through SocialSci and Qualtrics.
Participants viewed the same faces varying in perceived dominance and competence as in Experiment 1. Verbal prompts solely elicited “mean/nice” judgments (i.e., “Which of these people is very [mean/nice]?”). Unlike in Experiment 1, here Face Trait was a between-subjects variable. Sample sizes for all traits and age groups are displayed in Figures 1.5–1.7.

**Results**

As in Experiment 1, consensus of judgment was strikingly high, with consistency for all age groups and both traits vastly exceeding chance responding (50%), $ps < .001$, $ds > 2.15$. These results are summarized in Figures 1.5–1.6. Consensus of “mean/nice” judgments based on facial dominance (i.e., dominant = “mean”) showed developmental invariance, ranging from 87% to 95% with no significant pairwise differences between any age groups, all $ps > .06$, $ds < .50$. Children of all ages also showed robust consensus when attributing “mean/nice” to faces varied on competence (82%-96%). Adults, however, showed markedly lower attributions of “nice/mean” to the competent/incompetent faces (76%) than did 5-6 year olds (94%) and 7-10 year olds (96%), $ps < .01$, $ds > 1.03$. Consensus increased with age among children, with 7-10 year olds responding significantly more consistently than 3-4 year olds (82%), $p < .05$, $d = 0.83$. 
Figure 1.5. Average rates of attributing “nice” to submissive and “mean” to dominant faces (Experiment 2).

Figure 1.6. Average rates of attributing “nice” to competent and “mean” to incompetent faces.
Data were further analyzed using a 2 (Face Trait: dominance vs. competence) × 4 (age group: 3-4 vs. 5-6 vs. 7-10 year olds vs. adults) between-subjects ANOVA. Main effects emerged for both Age Group ($F(3,506) = 10.804, p < .001$) and Face Trait ($F(1,506) = 3.721, p = .054$), with a significant Age*Face Trait interaction ($F(3,506) = 2.674, p < .05$). When collapsing data across Face Trait, an age-related increase was observed among children, with 7-10 year olds showing significantly greater consensus than 3-4 year olds (Sidak post-hoc $p < .05$). Adults showed the same consensus as 3-4 year olds ($p > .99$) but were less consistent than 5-6 or 7-10 year olds ($ps < .01$).

Further analyses explored the possibility that participants might be more consistent when attributing “mean/nice” than when attributing the more specific traits of “strong/not strong” and “smart/not smart.” One-way ANOVAs using combined data from both studies compared consensus between “mean/nice” evaluations (Experiment 1) and specific trait judgments (Experiment 2) based on faces varying in dominance and competence. Overall, “nice/mean” judgments were significantly more consistent than judgments of “strong/not” and “smart/not” for faces varying in dominance ($F(1,629) = 5.332, p < .05$), and competence ($F(1,361) = 10.709, p < .01$), respectively. All groups of children were more significantly consistent when attributing “nice/mean” to faces as opposed to “strong/not strong” or “smart/not smart,” $ps < .05$. However, adults were less consistent when attributing “mean/nice” rather than “strong/not very strong” to faces varied on dominance ($F(1,362) = 6.441, p < .05$), and equally consistent when attributing “smart/not” and “nice/mean” to faces varied on competence ($F(1,133) = .816, n.s.$).
General Discussion

Children in both experiments made reliable inferences about character that approached adult levels at the earliest ages tested, and matched adult levels by age 7. In particular, participants of all ages robustly applied basic “mean/nice” judgments in response to a variety of facial characteristics.

In both experiments, judgments based on facial competence appeared to develop differently than those of trustworthiness and dominance. The data also produced the seemingly anomalous result that adults were less consistent than 7-10 year olds when attributing “nice/mean” to competence faces. It is possible that adult face-trait judgments might be more differentiated than those of children, who rely more on global valence. Heightened sensitivity with age to features other than those affecting global valence might also account for developmental increases in reliability when attributing strength and intelligence to faces (Experiment 1). The competence face model used here may also be less effective than those of trustworthiness and dominance.

The striking consensus in “mean/nice” attributions observed for all three trait dimensions suggests that such evaluations might underlie the consensus in face-trait inferences observed in Experiment 1. In fact, principal components analyses of trait judgments from faces show that trustworthiness judgments are strongly correlated with the first PC ($r > .90$), interpreted as valence (Oosterhof & Todorov, 2008), and the computational model of face trustworthiness closely resembles a valence model based on multiple social judgments (Said, Dotsch, & Todorov, 2010). Further research probing the
relationship between “mean/nice” judgments and specific trait inferences will be necessary to evaluate this possibility.

These two experiments provide a clear demonstration that children as young as 3-4 years of age show an adult-like tendency to attribute both traits and “mean/nice” evaluations to faces based on their appearance. It is possible that attractiveness underlies character inferences, particularly for trustworthiness and competence faces. However, recent work has shown that the facial features manipulated in these models elicit divergent trait judgments irrespective of attractiveness (Todorov et al., 2013).

These data leave open the question of when face-trait inference first emerges. Animation-based stimuli may enable researchers to study even younger populations. If such inferences take root early in development, as the data suggest, even infants might associate faces with trait-consistent behaviors, such as those conveying prosociality (Hamlin et al., 2007) or dominance (Mascaro & Csibra, 2012).

The predisposition to make rapid and unreflective judgments based on scant facial information is a pervasive form of social judgment. Prior work suggests that such inferences have important real-world consequences. We demonstrate that face-to-trait judgments are robust by age three, and certain judgments reach fully adult-like levels at 5-6 years of age. By revealing the young age at which children make such judgments, these data challenge accounts of slow-learning mechanisms of social learning that develop through the gradual detection and internalization of environmental regularities (e.g., Fazio et al., 1986; Smith & DeCoster, 2000).
Abstract

Humans rapidly and automatically use facial appearance to attribute personality traits ("trustworthy," "competent"). To what extent is this face-to-trait attribution learned gradually across development versus early in childhood? Here, we demonstrate that child-adult concordance occurs even when faces should minimize agreement: natural (not computer-generated) adult faces; less developed children’s faces; and perceptually unfamiliar monkey faces. In Study 1, 3- to 12-year-olds and adults selected “nice/mean” faces among pairs with a priori “nice-mean” ratings. Significant cross-age consensus emerged for all three face types. Study 2 replicated this result using an improved procedure in which 44–48 faces appeared in randomized pairs. This converging evidence supports the idea that complex forms of social cognition – allowing perceivers to believe they can derive personality from faces – emerge early in childhood, a finding that calls for new procedures to detect this central facet of cognition earlier in life.
Faces communicate a wealth of information about other people, and specialized mechanisms have evolved for the purpose of using faces to make inferences about others (Todorov et al., 2008; Zebrowitz, 1997). The past decade of research in social cognition has clearly demonstrated that not only do we decode basic group membership of individuals upon viewing their faces, but we also attribute personality traits to them based on their physiognomy. Whether accurate or not, judgments of character traits like trustworthiness, competence, and dominance are routinely and consensually offered. Prior work has shown that judgments about others’ personality from facial appearance form rapidly (Rule & Ambady, 2008a) and change little over time (Willis & Todorov, 2006). Importantly, face-based trait inferences are consistent both within and across cultures, and predict important outcomes such as election results (Ballew & Todorov, 2007; Todorov et al., 2005) and career success (Rule & Ambady, 2008b; Rule & Ambady, 2011). Face-to-trait judgments have thus emerged as a major source of interest for experimental psychologists seeking to understand the social cognitive bases of social perception.

We know a great deal about how adults attribute personality to others based on their physical appearance. For example, perceivers use static facial characteristics to inform decisions about personality traits and even complex beliefs such as religious affiliations (Rule, Garrett, & Ambady, 2010). However, little is known about the development of the fundamental tendency to infer traits such as “trustworthy” or “competent” from facial appearance alone. Understanding the developmental process can inform whether face-to-trait inferences represent a fundamental element of social cognition, in which case they should be observed relatively early in development.
Studying the developmental course of such judgments can also shed light onto the degree to which they require extensive social learning and experience both to initially emerge and to reach full maturity.

Previous work has shown that even infants possess the roots of face perception (Cassia, Turati, & Simion, 2004; Pascalis et al., 2008) and even demonstrate certain preferences based on facial attractiveness (e.g., Langlois et al., 2000) and group membership in domains like gender (Quinn et al., 2002), and ethnicity (Kelly et al., 2005; Bar-Haim et al., 2006). Recent work has also begun to suggest that face-to-trait inferences may emerge early in development. When asked to choose faces possessing certain traits, such as “nice,” “strong,” or “smart,” children as young as 3–4 years of age tended to choose the same types of faces as adults (Cogsdill et al., 2014). This cross-age consensus was especially pronounced when children were asked to make basic evaluative judgments (i.e., “Which of these people is very nice/mean?”), and all participants did so readily for all of the types of faces. Thus, the face perception skills and early social preferences that have been observed in infants also appear to manifest in face-based trait inferences that emerge by the preschool years, if not even earlier.

However, the research to date is limited in a variety of ways. Cogsdill et al. (2014) used only computer-generated face stimuli that were designed to strongly elicit judgments of high or low trustworthiness, dominance, and competence (Oosterhof & Todorov, 2008). Besides being computer-generated rather than natural in appearance, these faces were also specifically selected for their extreme representation of each trait. Natural faces representing a milder range of features may fail to elicit the same attributions from children.
To test the strength of face-to-trait inference across development, the present work measures cross-age consensus in large numbers of children and adults when evaluating naturalistic human faces. If children fail to produce adult-like responses while viewing natural human faces, this would suggest that Cogsdill et al. (2014) overstated the degree to which children form adult-like trait attributions in response to faces. If, on the other hand, the developmental continuity that was observed with more extreme computer-generated faces is mirrored here, this would provide compelling supporting evidence that child-adult agreement in face-to-trait inference can emerge in the natural world.

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

Yet another conservative test of the development of face-trait inferences can be achieved by using faces that are novel to both children and adults. Faces belonging to non-human primates provide such a set. Because participants are unfamiliar with monkey faces, these stimuli will reveal whether perceptual experience affects the development of consensual face-to-trait judgments. If we observe agreement between the judgments that children and adults apply to such unfamiliar stimuli, this would further corroborate the robustness of face-to-trait inferences even in early childhood.
Monkey faces confer an additional advantage in that they are less susceptible to the culture-specific trait inferences that may guide judgments of human faces. Even though the adult human faces we used were deliberately altered to appear relatively homogenous in appearance (through manipulations described below), some stereotypes may still influence judgments of people from different cultures. For example, even though all faces belong to European targets, some may belong to different groups within that ethnic designation, a fact that some perceivers might be able to recognize. The lack of culture-specific biases associated with monkey faces thus affords the ability to test the development of face-to-trait inference devoid of such influences.

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

**Experiment 1**

**Method**

**Participants**

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

*Stimuli*

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

The three types of faces differed significantly in their visual characteristics. Adult faces were obtained from a set of photos taken of emotionally neutral, White male actors in Karolinska, Sweden (Lundqvist, Flykt, & Ohman, 1998; Figure 2.1). All male photos in this set were cropped into a circular image such that participants were forced to rely only on internal facial features when rendering their judgments. Photos were converted to grayscale and image brightness was standardized across the set. These steps were taken to deliberately render the set of adult faces as homogeneous as possible, thereby further increasing the difficulty of the task in order to provide a more conservative test of correspondence between children’s and adults’ judgments.
Children’s faces were obtained from the Dartmouth Database of Children’s Faces (Dalrymple, Gomez, & Duchaine, 2013; Figure 2.2). While the full database includes faces of both males and females between the ages of 5 and 16, only a subset of 49 male faces was piloted for use in this study. The ages of children included in the final set of stimuli for Experiment 1 was more narrow than that of the full database, ranging from only 8 to 12 years of age ($M = 9.60, SD = 1.58$). As with adult faces, only male faces were included so as to avoid effects driven by gender stereotypes and to maintain consistency across the face conditions. Children were photographed wearing black shirts and caps such that their ears, hair, and clothing were not visible. Unlike the adult faces, however, some external features of children’s faces remained visible, and their faces were presented in full color with variable brightness.

Rhesus macaque faces were presented in the most straightforward manner, with monkeys’ entire heads simply being cropped out of larger images provided by Laurie Santos’ Comparative Cognition Laboratory (Figure 2.3). While only 20 faces were piloted for use in Experiment 1, an additional 24 were added to the set for Experiment 2. Faces were presented in full color in Experiment 1 and in grayscale for Experiment 2, and this did not appear to significantly alter participants’ responses.

Using the pilot ratings collected for each stimulus type, five “nice” and five “mean” faces were selected for use in each of the three face conditions (Adult, Child, and Monkey Faces). Faces were selected from across the full range of pilot ratings such that those ratings were matched across all three stimulus sets; i.e., for every face in a given set, we selected a face in both of the other sets whose average pilot rating closely
matched that face (pilot ratings for all faces are provided in Figure 2.4). This feature of the stimuli allows us to directly compare results obtained using the three face sets.

*Figure 2.1.* Examples of stimuli depicting adult faces, with “mean” faces (based on pilot ratings) in the top row and “nice” faces on the bottom row.

*Figure 2.2.* Stimuli depicting children’s faces, with “mean” faces in the top row and “nice” faces on the bottom row.
Figure 2.3. Examples of stimuli depicting rhesus macaque monkey faces, with “mean” faces in the top row and “nice” faces on the bottom row.

<table>
<thead>
<tr>
<th>Mean faces</th>
<th>Nice faces</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>M</em> = 2.42, 2.43, 2.42</td>
<td><em>M</em> = 3.42, 3.42, 3.38</td>
</tr>
<tr>
<td><em>SD</em> = 0.34, 0.32, 0.33</td>
<td><em>SD</em> = 0.19, 0.21, 0.19</td>
</tr>
<tr>
<td><em>M</em> = 2.11, 2.15, 2.08</td>
<td><em>M</em> = 3.19, 3.21, 3.19</td>
</tr>
<tr>
<td><em>M</em> = 2.24, 2.25, 2.28</td>
<td><em>M</em> = 3.27, 3.27, 3.24</td>
</tr>
<tr>
<td><em>M</em> = 2.26, 2.28, 2.32</td>
<td><em>M</em> = 3.43, 3.36, 3.34</td>
</tr>
<tr>
<td><em>M</em> = 2.53, 2.53, 2.48</td>
<td><em>M</em> = 3.55, 3.55, 3.48</td>
</tr>
<tr>
<td><em>M</em> = 2.96, 2.95, 2.96</td>
<td><em>M</em> = 3.66, 3.70, 3.67</td>
</tr>
</tbody>
</table>

*Figure 2.4. Pilot ratings of “nice” for all faces used in Experiment 1 (1 = meanest, 5 = nicest), based on independent ratings from *N* = 100 for each of the three sets of faces (total *N* = 300).*
Procedure

Children completed the experiment by pointing to a laptop while being seated facing the pairs of stimulus faces. They first completed a practice round consisting of three trials in which the instruction was simply to point to the larger or smaller face, a task that all children completed easily, ensuring that children understood the task and were pointing clearly to the correct face on the screen. No practice trials were included for the adult version, which participants completed on www.socialsci.com on their own computers. For the main experiment, participants viewed pairs of faces and answered the question, “which of these [people/children/monkeys] is very [nice/mean]?” Each face pair consisted of one “nice” and one “mean” face based on pilot ratings, with all combinations of five nice and five mean faces yielding a total of 25 trials. For all participants, the “nice” face appeared on the left side of the screen in either 12 or 13 of 25 trials. The order of face pairs was randomized such that no face appeared in consecutive trials, and prompts were randomized so that the anticipated response was on the left or right sides with approximately equal frequency (12 or 13 trials). All participants saw faces from only one stimulus set, making Face Type a between-subjects variable (Adult vs. Children vs. Monkey faces). Children were assured that there were no right or wrong answers.

Results

Overall patterns of response were quite consistent (Figure 2.5). Participants provided anticipated responses in 81% of all trials, a rate that was significantly higher than chance (50%; t(599) = 49.19, p < .001, d = 4.02). This consistency was similarly high when data

---

6 All data are publicly available through the Open Science Framework at the following URL: https://osf.io/5x9iq/
were separated by Age Group, with children and adults providing expected responses in 78% and 84% of all trials, respectively. Both children and adults were most consistent when judging monkey faces, for which they provided expected responses in 81% and 86% of trials respectively. Although adults were similarly consistent when judging faces belonging to children (84%) and other adults (82%), child participants were less reliable when judging those faces, giving anticipated responses in 80% of trials when viewing children’s faces and 72% of trials when viewing adult faces. All judgments by all age groups greatly exceeded chance responding at 50% (all \( t_{99} > 16.04 \), all \( p < .001 \), all \( d > 3.22 \)).

Data were first analyzed in the aggregate using a 3 × 2 ANOVA, with Face Condition as the first factor (Adult, Child, and Monkey faces) and Age Group as the second factor (Child vs. Adult participants). This analysis revealed a significant effect of Age Group, with adults providing more expected responses than children across the full set of data, \( F(1,594) = 25.74, p < .001, \eta^2 = .042 \). The effect of Face Condition was also significant, \( F(2,594) = 9.46, p < .001, \eta^2 = .031 \). The Age Group × Face Condition interaction was marginally significant, \( F(2,594) = 2.80, p = .062, \eta^2 = .009 \). Thus, adults and children differed in their responses, and the type of face being judged affected patterns of response, with the interaction between the two approaching significance. The significant main effects of Face Condition and Age Group were further analyzed. First, Sidak-corrected post-hoc analyses on the original ANOVA compared response patterns between the three Face Conditions, revealing that responses to adult faces (77%) were less reliable than those generated in response to faces belonging to children (82%, \( p < .01, d = 0.30 \)) and monkeys (84%, \( p < .001, d = 0.42 \)), which were not significantly
different from one another ($p = 0.68, d = 0.10$). Next, independent samples t-tests compared responses between adults and children in each of the three Face Conditions (Adult, Child, and Monkey Faces). Here, the largest difference in responses between adults and children was observed for Adult Faces, in which adults provided the expected responses in 82% of all trials, whereas children did so in only 72% of trials, $t(198) = 5.07, p < .001, d = 0.72$. In the Monkey Faces condition, adults provided the expected responses in 86% of all trials, while children did so in 81% of trials ($t(198) = 2.26, p < .05, d = 0.32$). Child-adult agreement was greatest in the Child Faces condition, in which adults provided expected responses in 83.9% and children in 80.1% of all trials, a difference that did not reach statistical significance ($t(198) = 1.64, n.s.$). The marginally significant Age Group × Face Condition interaction therefore resulted from the fact that cross-age agreement differed depending on the face being perceived, with the greatest consensus emerging in response to faces belonging to children.

The relatively large samples of children we collected enabled us to conduct correlational analyses to explore the developmental trajectories of judgments in the three Face Conditions. To do this, we measured Pearson correlations to study the effects of age on patterns of response among child participants, with adult participants’ responses being excluded from analysis. Here, the size of each correlation (i.e., the effect size) represents the magnitude of the effect of development on consensus of response. When data were collapsed across Face Condition ($N = 300$), the relationship between age and consensus was highly significant, $r = 0.30, p < .001$. All three Face Conditions yielded nearly identical effect sizes when analyzed separately (all $Ns = 100, rs = 0.33, ps < .001$; Figures
2.6–2.8), and further analyses using Fisher r-to-z transformations confirmed that the differences between these correlations were not significant (all $zs < 0.02$, all $ps > 0.98$).

To test for possible influences of outliers in these data, an informal analysis was conducted in which one outlier was removed from data collected for each of the three types of faces. Once these outliers were removed, the effects of age on consistency of response remained consistent for all three types of faces (changing to $r = .30$, .40, and .40 for monkey, adult, and child faces, respectively). These effect sizes were again not significantly different from one another.

*Figure 2.5.* Overall patterns of response from Experiment 1. Error bars represent 95% confidence intervals.
Figure 2.6. Scatter plot representing the relationship between age and consistency of response among child participants in Experiment 1 when judging monkey faces.

Figure 2.7. Age and consistency of response of child participants judging adult faces in Experiment 1.
Figure 2.8. Age and consistency of response of child participants judging other children’s faces in Experiment 1.

Discussion

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

Surprisingly, however, correlational analyses measuring the effects of development on consensus among children yielded practically identical effect sizes for the three types of stimuli used. This striking convergence further suggests the possibility that the significant interaction observed between Age Group × Face Condition may be due to developmental changes occurring during adolescence, during the ages between those of the oldest ages of children and the youngest adults we tested. The possibility that face-based evaluations might actually “mature” over a prolonged period of time over the course of development, even extending into adolescence, is consistent with recent work on face processing suggesting that face recognition (e.g., Lawrence et al., 2008) and that the neural regions implicated in face processing also develop in a highly prolonged fashion, reaching maturity in the adolescent years or later (e.g., Scherf et al., 2007). Such work suggests the possibility that the ability to consistently attribute traits to faces with subtle variations in appearance may follow a highly prolonged period of development as the full complement of face processing abilities matures.

This consistency of developmental effects across the three Face Type conditions clearly demonstrates that the type of face being judged was relatively unimportant in guiding the consistency of children’s responses. In other words, children’s judgments of faces appear to be sufficiently flexible to accommodate the range of physical features and visual characteristics that were present in the three types of faces that were used in Experiment 1. Thus, whatever intuitions children use to guide their face judgments appear to be present early in life, develop across childhood (and perhaps even into
adolescence), and are readily applied to diverse face stimuli, owing to their robust nature early in life.

**Experiment 2**

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

**Method**

*Stimuli*

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

Procedure

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

The experimental procedure was identical for all three Face Types, with the sole exception consisting of the number of trials presented, with 22, 24, and 23 trials of Adult, Child and Monkey faces, respectively. Children’s versions of the experiment began with the same three practice trials that were used in Experiment 1. During the main portion of the experiment, children were instructed to simply point to the “nice” face in each trial. The experimenter further explained that children could simply keep pointing to the nicer face in each pair when new faces would appear on the screen, rather than waiting for the experimenter to ask them which one was nicer. This was done to eliminate the need to repeat the same question in all 22–24 trials.
During each trial, a pair of faces would appear on the screen. In the first trial, children were prompted to “Start pointing to the nicer faces,” after which they would point to one of the faces on the screen. The experimenter would click on the face they selected, upon which the experiment would automatically advance to the next trial by displaying a new pair of faces. This continued for all 22–24 trials until all faces had appeared, at which time the child would be thanked and offered their choice of sticker.

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

Results

For each face in each stimulus set, the percentage of children and adults who selected the face as the “nice” one was recorded. Consensus was thus measured as the strength of the relationship between children’s and adult’s frequencies of categorizing individual faces as “nice.” These associations were measured using Pearson correlations between these frequencies for all faces in the three sets (Figures 2.9–2.11), which represent the degree of correspondence between children’s and adult’s selections of faces belonging to monkeys (N = 46 faces), adults (N = 44 faces), and children (N = 48 faces). These frequencies were obtained from samples of N = 100 adults and N = 100 children for each of the three face types (overall N = 600).
Figure 2.9. Scatter plot displaying the percentages of children and adults who selected each of the monkey faces as being “nice” in Experiment 2.

Figure 2.10. Percentages of children and adults who selected each of the adult faces as being “nice” in Experiment 2.
Figure 2.11. Percentages of children and adults who selected each of the children’s faces as being “nice” in Experiment 2.

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

Further analyses were conducted to observe how this child-adult consensus might develop with age among child participants. To do this, we systematically excluded children at each age (starting by excluding 13-year-olds, then 12-year-olds, and so on) and measured the correlations that result from comparing adult categorization frequencies to these progressively smaller and younger groups of children. This analysis allows us to investigate the lower limits of children’s tendency to generate adult-like face categorizations.
The results of this analysis, which are summarized in Tables 2.1–2.3, show that a significant correlation between children’s and adult’s face categorizations persists even when a large portion of the child samples are excluded from analysis. Specifically, for categorization of adult and child faces, correlations remained highly significant even when the samples were reduced to include only 3–5 year olds, with only \( n = 35 \) children viewing adult faces \( (r = 0.35, p < .01; \text{Table 2.1}) \) and \( n = 28 \) viewing children’s faces \( (r = 0.61, p < .001; \text{Table 2.2}) \). When viewing monkey faces, this relationship remained significant even when only the responses of 3–4 year olds were compared to those of adults \( (n = 18 \text{ children}, r = 0.45, p < .001; \text{Table 2.3}) \).

<table>
<thead>
<tr>
<th>Children's maximum age</th>
<th>N children</th>
<th>Child/adult correlation coefficients (N=44 faces)</th>
<th>t-value</th>
<th>p-value (df=46, two-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>100</td>
<td>0.81</td>
<td>9.02</td>
<td>0.00</td>
</tr>
<tr>
<td>11</td>
<td>98</td>
<td>0.81</td>
<td>9.08</td>
<td>0.00</td>
</tr>
<tr>
<td>10</td>
<td>94</td>
<td>0.81</td>
<td>9.04</td>
<td>0.00</td>
</tr>
<tr>
<td>9</td>
<td>87</td>
<td>0.80</td>
<td>8.58</td>
<td>0.00</td>
</tr>
<tr>
<td>8</td>
<td>81</td>
<td>0.76</td>
<td>7.70</td>
<td>0.00</td>
</tr>
<tr>
<td>7</td>
<td>74</td>
<td>0.76</td>
<td>7.70</td>
<td>0.00</td>
</tr>
<tr>
<td>6</td>
<td>53</td>
<td>0.62</td>
<td>5.12</td>
<td>0.00</td>
</tr>
<tr>
<td>5</td>
<td>35</td>
<td>0.44</td>
<td>3.13</td>
<td>0.00</td>
</tr>
<tr>
<td>4</td>
<td>19</td>
<td>0.18</td>
<td>1.20</td>
<td>0.24</td>
</tr>
</tbody>
</table>

Table 2.1. Additional analyses from Experiment 2 showing correlations between children’s and adults’ categorizations of adult faces with progressively younger samples of children.
This analysis shows that the consensus we observe between children’s and adult’s face categorizations emerges very early in development, with children as young as 3–5 years of age showing significant consensus when judging human faces and even faces of

Table 2.2. Correlations between children’s and adults’ categorizations of children’s faces in Experiment 2.

<table>
<thead>
<tr>
<th>Children's maximum age</th>
<th>N children</th>
<th>Child/adult correlation coefficients (N=48 faces)</th>
<th>t-value</th>
<th>p-value (df=46, two-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>100</td>
<td>0.78</td>
<td>8.40</td>
<td>0.00</td>
</tr>
<tr>
<td>12</td>
<td>97</td>
<td>0.78</td>
<td>8.36</td>
<td>0.00</td>
</tr>
<tr>
<td>11</td>
<td>93</td>
<td>0.78</td>
<td>8.40</td>
<td>0.00</td>
</tr>
<tr>
<td>10</td>
<td>92</td>
<td>0.78</td>
<td>8.49</td>
<td>0.00</td>
</tr>
<tr>
<td>9</td>
<td>82</td>
<td>0.77</td>
<td>8.20</td>
<td>0.00</td>
</tr>
<tr>
<td>8</td>
<td>68</td>
<td>0.78</td>
<td>8.46</td>
<td>0.00</td>
</tr>
<tr>
<td>7</td>
<td>54</td>
<td>0.75</td>
<td>7.76</td>
<td>0.00</td>
</tr>
<tr>
<td>6</td>
<td>44</td>
<td>0.68</td>
<td>6.32</td>
<td>0.00</td>
</tr>
<tr>
<td>5</td>
<td>28</td>
<td>0.61</td>
<td>5.22</td>
<td>0.00</td>
</tr>
<tr>
<td>4</td>
<td>13</td>
<td>0.20</td>
<td>1.36</td>
<td>0.18</td>
</tr>
</tbody>
</table>

Table 2.3. Correlations between children’s and adults’ categorizations of monkey faces in Experiment 2.
nonhuman primates. Additionally, since the set of children’s faces in Experiment 2 represented a much broader range of ages (5-14 years old) than those in Experiment 1 (8-12 years old), this provides an even clearer demonstration of the broad generalizability of face-to-trait inferences.

**General Discussion**

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

There are several possible reasons as to why this may be. One possibility is that participants simply become more consistent in their responding as they grow older, and that these results indicate a domain-general cognitive tendency towards consistency of judgment that should also apply to many different possible types of social judgments.
Another possibility is that stereotypes linking facial features to personality traits grow continually stronger through associations that are learned both explicitly and implicitly via social experience across the lifespan, and that this period of statistical learning is a prolonged one that extends through adolescence. Further research studying different types of social judgments across a similarly wide age range will help to explore the reasons underlying the apparent prolonged development of judgments observed here.

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

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

In addition, the finding in Experiment 1 that the difference in consistency between adults and children was greater when attributing “nice/mean” to adult human faces as opposed to monkey faces suggests that developmental changes in perceiving traits in human faces might be driven by a gradual accumulation of perceptual experience with those faces, as children encounter human faces in everyday social environment with much greater frequency than monkey faces. At the same time, however, the fact that child-adult consistency was similar across the three types of faces in Experiment 2 suggests that while this perceptual experience could potentially play a larger role with human faces, the magnitude of its role in development is relatively small. Nevertheless, future research would benefit from studying older adolescent participants (ages 13–17) in order to assess whether developmental changes do indeed persist throughout adolescence, or whether developmental changes cease to be observed at a certain point. Another useful avenue for future work would be to directly measure the effects of perceptual experience by manipulating participants’ exposure to unfamiliar face stimuli. If perceptual experience does influence the frequency with which children dispatch adult-like judgments to faces, this would suggest a possible mechanism by which judgments continue to become increasingly consistent throughout adolescence and into adulthood.

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

Cogsdill, E. J., Spelke, E. S., & Banaji, M. R.

Abstract

Humans possess a strong tendency to infer character traits such as trustworthiness and dominance from facial appearance. Recent work has shown that even children as young as 3-4 years of age render adult-like character assessments, such as “nice/mean,” “strong,” or “smart,” in response to a variety of face stimuli. While such work has shown that basic evaluations are highly robust, it remains unknown to what degree children’s trait inferences generalize to other social judgments and behaviors based on faces. In Experiment 1, children attributed both “nice/mean” evaluations and trait-relevant behaviors to faces previously identified as “nice/mean,” “strong/not” and “smart/not,” showing that young children’s trait judgments generalize to behavioral predictions. The experiments that follow show that children are also more likely to prefer novel objects endorsed by “nicer”-looking faces (Experiment 2), and they are also more likely to give gifts to those faces (Experiments 3-5). In addition, Experiment 5 showed that children’s evaluations of which faces were “nice” directly predicted their likelihood of giving a gift to them. Children thus readily attribute behaviors and even modulate their own behaviors towards others based on the facial appearance of others, demonstrating that face-to-trait inferences are pervasive and consequential from the earliest years in life.
Forming impressions about the character and personality of others lies at the core of human interaction, and the face is used as a rich source of information. From past research we know that within the first moments of viewing a face, humans automatically form a number of evaluations. Mere milliseconds of exposure to static, two-dimensional face images enables perceivers to detect membership in group categories like gender, age, and race (Cunningham et al., 2004; Ito, Thompson, & Cacioppo, 2004; Valdés-Conroy et al., 2014), and even to form impressions of underlying qualities of character traits like trustworthiness, dominance and competence (for a review, see Todorov, Said, & Verosky, 2011). Such judgments, irrespective of their accuracy, are known to form within 100 milliseconds after exposure (Willis & Todorov, 2006), and evidence from functional neuroimaging has suggested that amygdala activity to faces correlates with consensual judgments of facial trustworthiness (Engell, Haxby, & Todorov, 2007). Face-to-trait inference is therefore not only rapid and unreflective but even automatic in nature. Moreover, character judgments based on facial appearance have been shown to be consequential in a variety of ways, such as by predicting voting behavior (Todorov, Mandisodza, Goren, & Hall, 2005; Hall, Goren, Chaiken, & Todorov, 2009), military rank (Mueller & Mazur, 1996), and harshness of criminal sentencing (Blair, Judd, & Chapleau, 2004).

As such, face-to-trait inferences represent a form of social cognition whose effects have widespread consequences throughout society, and recent work has shown that these judgments may be sufficiently fundamental that they are observed even in early childhood (Cogsdill, Todorov, Spelke, & Banaji, 2014). Furthermore, this ability is quite robust, as similar cross-age agreement emerges even in response to three different types
of naturalistic faces that were selected to appear more subtle in appearance, with some faces even belonging to non-human primates (Cogsdill & Banaji, 2014). In addition to rendering judgments based on static facial structure, children are also sensitive to facial expressivity, and provide reduced ratings of “friendliness” in response to short video clips of other children with autism spectrum disorder (Stagg, Slavny, Hand, Cardoso, & Smith, 2014). Such work clearly demonstrates that, like adults, children consistently render a number of deeply social judgments based on facial appearance alone.

One could argue, however, that face-to-trait inferences are a more meaningful form of social cognition if they inspire other social judgments and behaviors. The wealth of research on adult face judgments, while intriguing in its own right, is of particular significance because of the impressive downstream outcomes that such judgments have been shown to influence. While new research showing that children form character inferences from faces in adult-like patterns suggests that similar processes emerge early in development, it remains to be demonstrated that such early forms of social cognition are sufficiently robust to produce other social judgments and behaviors that combine to shape the social world of the developing child.

The existing research on early-forming impressions based on facial appearances leaves open the possibility that although children can attribute traits to faces, and even do so with adult-like consensus, such judgments may not generalize to behavioral attributions consistent with such judgments. This question is therefore the topic of the present investigation. If children’s face-based character judgments also translate to attributions of behavior as well as children’s own behaviors towards others possessing those facial features, such a set of findings would reveal the existence of a measurable
connection between mental representations of others and resulting manifest behaviors, which would underscore the importance of face-based character inference above and beyond past work showing that such judgments are visible early in childhood. It would also demonstrate that children’s early face-to-trait inferences are, in and of themselves, consequential in the social life of the child. If children attribute behaviors and modulate their own behaviors towards others based on facial appearance alone, this could influence how they choose whom to trust, learn from, and affiliate with from the earliest years of life. If supported, this possibility would have ramifications for how scientists think of the role that children play in their own socialization and even their own social education. This is especially important in light of the growing recognition that young children actively decide whom to consult for information and help solving problems (e.g., Baldwin & Moses, 1996; Cluver, Heyman, & Carver, 2013; Harris, 2012), even using physical appearance to decide who to learn from (Bascandziev & Harris, 2014), as well as the broader importance of social learning as an adaptive evolutionary mechanism (e.g., Tomasello, 1999; Williamson & Markman, 2006).

The present work aims to explore two specific questions regarding the consequences of face-to-trait inferences that emerge early in development. The first question is, to what degree do children’s trait judgments based on faces result in their expectations of trait-relevant behaviors? We investigate this in Experiment 1 by asking children and adults to use two-dimensional static representations of human faces to render character judgments via traits, such as “nice” or “mean,” “strong,” or “smart,” as well as behaviors signifying trust such as “[helping] other people when they are in trouble” or behaviors signifying competence such as “[drawing] pictures that look just
like real life” (see Table 3.1 for a full list of behaviors). The second question concerns whether children modulate their behaviors towards others in response to perceived character traits they extract from facial features. We explore this question using two general approaches: first, by asking children to select which item they would prefer to receive after having heard about the objects that two novel characters prefer (Experiment 2); and, second, by asking them to give a small gift, such as a token representing a cookie, to one among two individuals represented only by a face image (Experiments 3-5). These experiments test the possibility that children might behave differently towards people in a manner consistent with the traits that they attribute to them based on facial appearance alone.

All experiments reported here use an uncommonly large sample for such studies, with a wide age range consisting of adults as well as children between the ages of 3 and 10, in some cases up to age 13. This broad diversity of ages allows for the analysis of social judgments from toddlerhood to adulthood and has the potential to generate insights about the trajectory of social cognitive development.

**Experiment 1: From Trait Inferences to Behavior Attributions**

Experiment 1 probes the question of whether early face-based character inferences also lead to attributions of trait-relevant behaviors. Previous research studying the developmental origins of trait understanding has generally suggested that young children (i.e., around 5 years of age) possess only a tenuous understanding of behavioral consistency, and therefore personality traits in general. Such work has typically shown that children’s ability to use past behaviors to predict future behavior is one that only
emerges readily under the right testing conditions. For example, when children are asked to make a dichotomous choice between two possible future behaviors, they do not appear to show such an understanding of behavioral consistency until around 7 years of age (Rholes & Ruble, 1984, 1986; Rotenberg, 1980, 1982; Ruble, Newman, Rhose, & Altshuler, 1988). However, if children are asked to predict behavioral consistency in certain specific ways, they are able to do so from an early age. For example, if they are asked to make a more quantitative assessment reflecting their judgments of what behaviors are likely given past behavior, for example by asking children to estimate the number of pennies a character would share with them (e.g., Anderson & Butzin, 1978; Dozier, 1991) even five-year-olds will demonstrate signs of this early form of personality understanding. Although this and other examples (e.g., Boseovski & Lee, 2006; Cain, Heyman, & Walker, 1997) show that children may be capable of understanding behavioral consistency, they are generally more reluctant to do so than adults.

The fact that children in prior work readily make character assessments in response to faces raises the intriguing possibility that although children may not readily use prior behaviors to guide their predictions of future behaviors, they may in fact do so using a stimulus that is as superficial as a two-dimensional, static, computer-generated face image. In other words, although information about prior behaviors might be expected to be a stronger signal of future behaviors than mere facial appearance, it is possible that young children may actually use this seemingly impoverished stimulus to render such predictions. Consistent with this possibility, prior work has shown that although children struggle with behavior-to-behavior prediction tasks, they readily go from behaviors to traits and from traits to behaviors when the two types of questions are
presented separately (Liu, Gelman, & Wellman, 2007). Given that prior work has shown children can go from a face to a trait (as in Cogsdill et al., 2014), and from a trait to a behavior (as in Liu et al., 2007), can children go directly from a facial image to a prediction about future behavior? If true, this would suggest that children’s earliest applications of their nascent understanding of personality traits may be based in face-based trait inferences applied to unfamiliar others, even in the absence of any information about past behavior or character.

The goal of Experiment 1 was thus to clarify the extent to which children possess the intuition that facial features are indicative of not only character traits (like “nice”) that indicate, broadly speaking, whether a stranger should be approached or avoided, but also whether these evaluations of global valence lead in turn to predictions of specific behavioral tendencies such as helping others and sharing food. Whereas simple trait inferences capture globalized valence judgments, which perceivers can use for snap judgments as to whether an unfamiliar person is safe to approach or avoid, specific behavioral predictions as a result of these initial impressions would suggest that children are capable of stereotyping others based on facial information alone to an even more explicit degree than is expressed in global “nice/mean” evaluations, If this experiment shows that adults but not children infer behaviors from faces, we will know that this stage of inference requires protracted developmental emergence. If, on the other hand, children and adults equally infer behavioral tendencies from static two-dimensional faces, and at an age similar to previously demonstrated ages for face-trait inference, such a result would reveal that (a) information about both dispositions and behaviors are interwoven early in development, and that (b) trait and behavioral inferences of the sort that adults
readily produce are formed at the same early stage in childhood. As such, it would uncover the extent to which children’s early-forming understanding of personality includes the adult-like intuition about physical facial features as indicative of underlying character, tenuous as their predictive validity may be. The latter result would also indicate that children have a more robust understanding about character traits than was previously thought.

**Method**

**Stimuli**

Experiment 1 used computer-generated face stimuli from publicly available sets created by Oosterhof & Todorov (2008). As in Cogsdill et al. (2014), faces represented the high and low ends of the dimensions of trustworthiness, dominance, and competence. In fact, the exact same set of 18 faces that appeared in Cogsdill et al. (2014) were used again in Experiment 1, with three each of trustworthy/untrustworthy, dominant/submissive, and competent/incompetent faces. Examples of both Extreme and Moderate face stimuli are provided in Figure 3.1.
Design

The experimental design was identical for children and adults, a unique and beneficial feature of these types of experiments. The experiment proceeded in two blocks of 12 trials each, with one consisting of “nice/mean” questions and the other consisting of behavioral attribution questions. The order in which these two blocks appeared was counterbalanced across all participants. Within the Nice/Mean block, children were asked to identify the nice or mean face in each pair a total of twelve times (“Which of these people is very [nice]?”), and the order in which all six “nice” and “mean” questions were asked was randomized for each participant. Within the behavior attribution block, two behaviors probing each of the three traits were asked twice, again for a total of 12 trials (Table 3.1). As in the “nice/mean” block, behavioral questions were posed in unique randomized order for each participant.
Table 3.1. A list of all behavioral items used in Experiment 1.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Behavior (“Which of these people ____”)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trustworthiness</td>
<td>“helps other people when they are in trouble?” “likes to share their food with other people?”</td>
</tr>
<tr>
<td>Dominance</td>
<td>“can pick up really heavy things?” “always decides which game to play?”</td>
</tr>
<tr>
<td>Competence</td>
<td>“knows how to sing a lot of different songs?” “can draw pictures that look just like real life?”</td>
</tr>
</tbody>
</table>

Throughout both blocks, three types of trials corresponded to the three trait dimensions being studied: trustworthiness, dominance, and competence. In each of these three types of trials, the faces that appeared were high and low on the trait dimension, and, in the Behavior Attribution block, the question being asked pertained specifically to that trait. For example, one of the Trustworthiness behavior question would be, “Which of these people likes to share their food with other people?”, whereas a Competence trial question would be “Which of these people knows how to sing a lot of different songs?”. For each participant, the three types of trials proceeded in one of six possible fixed orders (e.g., Trustworthiness-Dominance-Competence), which were counterbalanced across all participants.

In addition to extending previous research to understand face-to-behavior inferences, Experiment 1 also probed the sensitivity of both trait and behavior judgments in response to variations in stimulus extremity. This was accomplished by introducing
Face Extremity as a between-subjects variable, whereby approximately half of all participants viewed faces that were more strikingly varied along trait dimensions (+/- 3 standard deviations from neutral appearance), referred to as the “Extreme” face condition, while the other half in the “Moderate” condition viewed the exact same face identities with a more subtle alteration along the same personality dimensions (+/- 1 standard deviation). This manipulation represented an early exploration into whether the consensus of face-to-trait inferences we observe early in life is sufficiently robust to generalize to stimuli that are more subtle in appearance, a question that was explored in greater depth in Cogsdill & Banaji (2014). Unlike the stimuli in Cogsdill & Banaji (2014), however, which used entirely different categories of faces to create a more difficult task (adults vs. children vs. monkey faces), here this was achieved using stimuli that were nearly identical except for subtle differences in facial features.

The results from Experiment 1 can speak to the robustness of early face-to-trait inference in two ways: first and foremost, across judgment tasks, from trait inferences to behavior attributions; and second, across different types of stimuli, with one type appearing more extreme and the other more subtle in the traits they were designed to convey.

Procedure
A total of $N = 99$ children, who ranged in age from 3.17 to 10.5 years of age ($M_{age} = 6.46$ years, $SD = 1.92$), participated in person at the Boston Children’s Museum. Participants were mostly White (47%), Asian (16%), and Hispanic (12%), with the remaining participants being either identified as other races (10%) or declining to specify
(14%). Roughly equal numbers of participants were female (57%) and male (43%). In addition, $N = 50$ adult participants ranging in age from 19 to 66 years ($M_{age} = 32.46, SD = 11.13$) were recruited through Amazon Mechanical Turk and completed the experiment through www.socialsci.com. The majority of adults were White (76%), and genders were equally represented, with around half identifying as female (52%).

Children viewed pairs of faces on a laptop computer, and were guided through three practice trials to both acclimate them to the task and to confirm that the child could clearly select face stimuli via finger pointing. In the practice trials, children were asked to point to the larger or smaller of two faces. During the main experiment, children viewed pairs of faces sequentially. During each trial, the experimenter would ask a question – for example, “Which of these people is very nice?” in a Nice/Mean trial, or “Which of these people helps other people when they are in trouble?” in a Behavior trial. Children would respond by pointing to one of the two faces on the screen. At the conclusion of the experiment, children were offered their choice of a sticker and their adult family members were verbally debriefed.

Adults completed an online version of the experiment that was nearly identical to the one given to children, with a couple of minor differences. First, no practice trials were given prior to the start of the experiment. Additionally, whereas 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 in which faces and prompts appeared to adults was completely randomized. Thus, although most participants had roughly equal numbers of trials with expected responses on the left and right sides of the screen, random chance likely caused these numbers to be particularly skewed for some participants.

**Results**
The dependent variable of interest in Experiment 1 was the percentage of trials in which participants selected the face that was intended to go with each trait word, or trait-related behavior. In each trial, participants selected either the face that was designed to elicit the accompanying judgment (e.g., a trustworthy face in response to “Which of these people is very nice?”), or the face that was not designed to elicit the judgment (e.g., an untrustworthy face in the same example). Thus, following the same design used in Cogsdill et al. (2014), the rate at which participants made “expected” rather than “unexpected” responses was taken as a measure of consensus in face-based character assessments, with higher numbers indicating an adult-like tendency to consistently select certain faces in response to questions regarding personality traits.

Participants were tested in either the Extreme face condition, in which face stimuli were +/- 3 standard deviations from average faces for all three trait dimensions; or the Moderate face condition, in which the faces were less extreme in appearance (+/- 1 SD). Both conditions used the same face identities, with the extremity of the trait manipulations being the only difference between the two sets of stimuli. As such, data were analyzed separately for participants in the two groups.

“Extreme” Stimuli Condition

Adults and children were overall highly consistent in their response patterns (Figure 3.2). Unsurprisingly, 3-4 year olds were the least consistent responders, giving expected responses in 72% of all trials. Older groups of participants were progressively more consistent, with 5-6 year olds generating the expected trait in response to each face in 81% of trials and 7-10 year olds in 89% of trials. Adult consistency was at 81%, a rate
that was most similar to that of 5-6 year old children. Overall responses from all age groups were significantly greater than chance (50%; all ps < .001).

Data were further analyzed using a $4 \times 3 \times 2$ mixed factorial ANOVA with Age Group as a between-subjects factor (children ages 3-4, 5-6, 7-10, and Adults), and Trait (trustworthiness, dominance, and competence) and Block (nice/mean vs. behavior) as within-subjects factors. This analysis revealed main effects for all three variables. First, there was a main effect for Block, owing to the fact that nice/mean trials elicited more consistent responses than Behavior trials ($F(1,145) = 101.50, p < .001, \eta^2 = 0.41$; see Figure 3.3). A main effect also emerged for Trait, $F(1.88,272.85) = 26.57, p < .001, \eta^2 = 0.16$ (see Figure 3.4), with pairwise comparisons between the three traits showing that

---

8 Throughout the analyses for Experiment 1, we find that adults are often less consistent than 7-10 year old children, sometimes even significantly so. These counterintuitive differences may be partly due to the way in which the experiment was conducted. Adults participating online might have completed the experiment more quickly and with reduced attention compared to children. It is also possible that adults develop more nuanced and/or idiosyncratic judgments that may introduce some additional variability.

9 Greenhouse-Geisser corrections were applied whenever data failed Mauchly’s test of sphericity, $p < .05$. 

---

Figure 3.2. Overall results across all trials for all age groups in the Extreme faces condition in Experiment 1. Error bars in all graphs represent standard error.

Figure 3.3. Overall results across all trials for all age groups in the Extreme faces condition in Experiment 1. Error bars in all graphs represent standard error.

---

71
Trustworthiness trials (89%) were responded to more consistently than Dominance (80%, $p < .001$), which was in turn more consistent than responses for Competence trials (75%; $p < .05$). This almost perfectly replicates the pattern of results that was originally obtained in Cogsdill et al. (2014), with consensus obtained for the three types of faces being highest for those varying in trustworthiness (92%), followed by dominance (82%) and competence (74%).

Finally, as in prior studies, Age Group had a significant main effect, $F(3,145) = 8.42, p < .001, \eta^2 = 0.15$. Sidak-corrected post-hoc tests showed that this main effect was borne out by a significant difference between 3-4 year olds and all older groups of participants (all $ps < .05$), with no other pairwise contrasts between age groups reaching significance ($ps > .08$). Owing yet again to the striking convergence in results across studies, this pattern replicates data from Cogsdill et al. (2014), in which the same pattern of general age differences was observed for faces varying in trustworthiness and dominance, though not competence. Thus, the overall trends that emerge from the foregoing analyses are that although all types of judgments were highly reliable among all age groups and trait types, judgments were particularly consistent when faces varied in trustworthiness and when participants were at least 5 years of age.
Further analyses were conducted to specifically investigate the question of whether trait inferences generalize to behavior attributions. Although the preceding analyses showed that consensus was overall lower for behavior attributions compared to
nice/mee judgments, the interaction between Block × Trait was also significant
\(F(1.69,245.18) = 9.13, p < .001, \eta^2 = .059\), suggesting that the generalizability of
consensus from trait to behavior may have varied considerably between trustworthiness,
dominance, and competence trials. To explore this possibility, paired samples t-tests were
conducted to test the effect of Block – i.e., the relative decrease in consensus when going
from traits to behaviors – separately for each of the three Traits. This analysis showed
that although the effect of Block was significant for all three traits (all ts > 4.01, all ps <
.001, all ds > 0.46) the effect was most pronounced for Dominance trials. Here, the
difference in consensus was greatest, at 19.6% (\(t = 7.31\)), than that observed for both
Competence (10.0%, \(t = 4.36\)) and Trustworthiness (7.9% difference, \(t = 4.01\)). Since the
decrease in consistency between attributing traits and behaviors was most pronounced for
Dominance, this suggests that the link between trait inferences and behavioral attributions
was weakest for that trait. Thus, the significant Block × Trait interaction reveals that
although children and adults readily use trait inferences to guide behavioral attributions
based on perceived trustworthiness and competence, which saw relatively little decrease
in consensus between trait and behavior trials, the link between traits and behaviors may
be more tenuous in the domain of dominance.

It is possible that this interaction is due to a disconnect between the trait inference
requested (“strong” or “not strong”) and the behaviors that were used to probe
dominance. We would speculate that this may be because for trust and competence a
simpler valence orientation exists: high trust is better than low trust and high competence
is better than low competence. For dominance, however, this underlying valence
dimension is more complex. Not only do there exist more varied definitions of
“dominance,” which can encompass such differing capacities as physical strength (lifting heavy things) and social influence (deciding which game to play), but dominance also varies in valence depending on coalitions and whether the dominant individual belongs to the perceiver’s ingroup or not. Thus, the discrepancy we observed between dominance and the other two traits, trustworthiness and competence, may primarily result from differences in the complexity of evaluating others along those dimensions.

In addition to replicating multiple findings from prior research, the present results clearly demonstrate that children attribute trait-relevant behaviors to faces manipulated on those dimensions, and further that this convergence in judgments was most strongly pronounced in the domains of trustworthiness and competence. Thus, even from a very early age, character assessments based on faces extend into the realm of behavioral attributions, further supporting not only the early emergence but also the broad social relevance of face-to-trait inferences in social cognitive development.

"Moderate" Stimuli Condition
As explained above, all elements of the experimental design and procedure were identical in the “Moderate” faces condition, with the sole exception being that face stimuli were manipulated to a less extreme degree (+/- 1 SD from neutral) as compared with their counterparts in the Extreme faces condition. As such, the same steps were taken to analyze data resulting from the use of these stimuli.

A total of $N = 107$ children between the ages of 3.16 and 10.83 years of age ($M_{age} = 6.50, SD = 1.92$) completed the Moderate faces condition. Participants were mostly White (63%) and genders were represented equally (55% female). As in the Extreme faces condition, a separate group of $N = 50$ adults participated online, a sample which was also mostly White (74%) and about half female (62%).
Consistent with the use of more subtle stimuli, participants of all age groups responded with reduced consensus overall. The youngest children, ages 3-4 years, were yet again the least consistent responders, pointing to the expected face in 62% of all trials on average. 7-10 year olds did so in 76% of all trials, making them the most consistent age group, while 5-6 year olds and adults yielded intermediate levels of consensus at 71% and 75% respectively (see Figure 3.5). Although participants were less consistent in their tendency to select the expected face in every trial compared to those who viewed Extreme faces, all were still above chance when collapsing across Trait (Trustworthiness, Dominance, and Competence) and Block (Nice/Mean and Behaviors). When data were separated by the within-subjects variables of Trait and Block, two instances emerged in which consensus did not exceed chance. Specifically, when attributing behaviors to faces, the youngest group of children (ages 3-4) were at chance for trustworthiness and dominance trials ($p_s > .18$), although they were significantly above chance when responding to competence trials ($p < .01$). All other combinations of Age Group, Trait, and Block were significantly above chance, $t_s > 2.16$, $p_s < .05$, $d_s > 0.48$. 
Figure 3.5. Overall results for Extreme vs. Moderate faces conditions in Experiment 1. Symbols indicate two-tailed significance of difference between face conditions: * $p < .05$, ** $p < .01$, *** $p < .001$.

As was previously done with data from the Extreme faces condition, a 4 × 3 × 2 Repeated Measures ANOVA was conducted, with main effects emerging again for each of the three variables: Age Group (between-subjects; $F(3, 153) = 9.84, p < .001, \eta^2 = 0.16$), Trait (within-subjects), and Block (within-subjects). As can be expected following the results from the Extreme faces condition, participants were more consistent when responding to Nice/Mean as opposed to Behavior questions, which resulted in a main effect of Block ($F(1, 153) = 38.2, p < .001, \eta^2 = .198$). The variable Trait also yielded a main effect ($F(2, 306) = 9.69, p < .001, \eta^2 = .058$), with subsequent pairwise comparisons showing that Trustworthiness (72.8%) and Competence trials (73.3%) both yielded significantly higher consensus than Dominance trials (66.2%; $ps < .001$), yet were not significantly different from one another. This marks a departure from the results of the
Extreme faces condition, in which Trustworthiness, Dominance, and Competence yielded successively and significantly lower degrees of consensus. This likely reflects the fact that the decrease in consensus between the Extreme and Moderate face conditions was highly significant for faces varying in Trustworthiness (89% vs 75%; \( t(304) = 7.14, p < .001, d = 0.82 \)) and Dominance (80% vs. 67%; \( t(304) = 6.42, p < .001, d = 0.74 \)), while Competence was not significantly affected by the change in stimuli (75.2% vs. 74.5%; n.s.).

The preceding analyses also yielded a marginally significant Block × Trait interaction (\( F(2, 306) = 3.01, p = 0.05, \eta^2 = .019 \)) suggesting that, as was previously observed for Extreme faces, patterns of responses across the three different traits may have differed for Nice/Mean and Behavior attribution questions. Paired t-tests comparing consensus between these two types of questions revealed the same pattern observed with Extreme faces, namely that the decrease in consistency when attributing behaviors compared to “nice/mean” was most pronounced for faces varying in Dominance (\( t(156) = 4.91, p < .001, d = 0.56 \)), with progressively weaker differences observed for Trustworthiness (\( t(156) = 3.66, p < .001, d = 0.41 \)) and Competence trials (\( t(156) = 2.62, p < .05, d = 0.30 \)). This reflects the same issue, noted previously, that dominance may provide unique difficulty in transferring trait inferences to behavioral predictions.

**Discussion**

The results from both the Extreme and Moderate face conditions clearly show that children not only attribute traits to faces, including basic “nice/mean” evaluations, but they also attribute trait-relevant behaviors to those same faces. In addition to providing clear and consistent replications of prior research, these data also generate novel insights
regarding perceivers’ ability and willingness to use facial features to guide behavioral expectations. In particular, across the different types of faces used – whether they were designed to vary in trustworthiness, dominance, or competence, and whether they were varied “moderately” or “extremely” along personality dimensions – even the 3-4 year old children consistently used facial features to guide their judgments about how others are likely to behave.

The methodology of this experiment also highlights the robustness of children’s early face-based social judgments by showing the diversity of circumstances under which they can be reliably measured. Whereas in Cogsdill & Banaji (2014) this convergence was demonstrated by using dramatically different types of faces representing different ages and species, here we show a striking convergence in judgment across age groups using different types of questions, namely by using ones that focus on behaviors rather than traits. While this distinction may seem subtle, it is important to establish that the judgments children make about others are socially meaningful. By showing that they attribute both traits and behaviors to faces, we show that even very young children are capable of connecting outward appearance to behavioral tendencies, and such judgments are likely to play a role in shaping social experiences and relationships early in life.

The responses that emerge between the two face conditions, which used Extreme and Moderate face stimuli, reveal both remarkable consistencies and intriguing differences. Although overall consensus was lower among participants who viewed Moderate faces, as we should expect to find when participants view stimuli with features that are less conspicuously manipulated, responses were generally quite consistent across all ages, traits, and question types. The fact that most of the major findings were nearly
identical across the two conditions suggests that the early-emerging consensus in social judgments, which children and adults show when applying both traits and behaviors to faces, in fact reflects a social cognitive tendency that is highly stable across the lifespan. Perhaps the most intriguing difference that emerged among participants tested with Moderate versus Extreme faces was the preservation of consensus for faces varying in competence – i.e., the relatively stable consensus that emerged between participants who viewed faces that were more or less dramatic in appearance. Whereas judgments of faces varying in trustworthiness and dominance appear to be quite sensitive to the strength of the manipulations of facial features, judgments of faces varying in competence appeared unaffected by the degree of visual differences that differentiated the two stimulus conditions. This may suggest a different underlying mechanism of competence perception, perhaps one that is more categorical than continuous in nature. This possibility is only speculative, however, and remains to be evaluated by specific tests in future research.

Throughout the various nuances in the results outlined above, a clear picture emerges. First and foremost, early-emerging character inferences from faces do also extend to attributions of trait-relevant behaviors. Regardless of whatever difficulties children may have in understanding behavioral consistency and enduring personality traits, the present work shows that they are perfectly willing to use mere facial appearance as a cue to understanding how others are likely to behave. In addition, data from faces that were more subtle in appearance show that this ability to go from a face to both a trait and a behavior is both sensitive to degrees of variation in physical features, yet also highly robust, even at the earliest ages tested.
Experiments 2-5: Face-Based Modulation of Prosocial Behaviors

The experiments that follow investigate a related question by asking whether, like their adult counterparts, children actually behave differently based on others’ facial appearances. One possibility is that although adults show clear differentiations in behavior based on facial appearance, children might not act on their own face-based trait inferences in any meaningful way. It may well be that children’s character judgments are only elicited when they are directly asked to make such judgments. If children do not generate these inferences spontaneously and without needing this kind of explicit prompting, they would be unlikely to behave differently towards others as a result of differences in appearance. If, however, children show significant behavioral differences in response to their social judgments based on others’ facial structures, this would suggest that character judgments based on these appearances exert a powerful influence on social development throughout childhood, with the potential for direct consequences on how children interact with others in their early social environments.

Prior work studying adults has shown that face-based character judgments relate to not only important outcomes like criminal sentencing and career success, but also to more immediate behaviors such as those signaling trust. Recent work has shown that first impressions of trustworthiness based on facial appearance influences the initial actions taken by participants playing a repeated Trust Game, such that participants were more likely to invest in other players who they judge to appear trustworthy (e.g., Campellone & Kring, 2013; Chang et al., 2010; Van ’t Wout and Sanfey, 2008), although such judgments may ultimately be overridden by the reciprocating (or non-reciprocating)
actions that these opponents make (Yu, Saleem, & Gonzalez, 2014). This work has only recently begun to study the development of such tendencies in childhood, with one study finding that 5- and 10-year-old children show a similar preference for investing in trustworthy-looking others using a modified version of the Trust Game (Ewing, Caulfield, Read, & Rhodes, 2014). Moreover, these children were actually willing to invest an additional amount of money to gain visual access to their partners’ faces in the game. The fact that children willingly spent investment capital within the context of these games to view their partners prior to deciding whether to trust them or not suggests that they consider appearance cues to be economically beneficial, specifically by affording more accurate predictions of whether others are likely to be trustworthy or not. This information in turn causes children to calibrate their own behaviors accordingly.

Research directly measuring the effects of facial appearance on children’s behaviors is scant. Nevertheless, related bodies of work can provide clues to understanding how the link from seeing someone to behaving towards them could manifest early in life. One behavior that has been extensively studied is that of infants’ looking patterns, which are often interpreted as revealing social preferences. Such work has shown that, for example, infants prefer to look at faces that are more attractive (Langlois et al., 1987) or are the same sex as their primary caregiver (Quinn et al., 2002). Infants as young as only 3 months also show preferences for others that share their ethnicity (Bar-Haim, Ziv, Lamy, & Hodes, 2006). Looking-time based measures may provide a means of studying modulation of social behaviors in response to dimensions of facial appearance beyond group membership and attractiveness even in infancy.
Although such methods may one day enable researchers to infants’ social preferences based on facial structures, the focus of the present investigation is on concrete actions performed by children ages three and up. With this age range in mind, past research that has studied the behaviors of preschool age children can inform an approach to studying how facial appearance might modulate children’s behaviors. One example of such work is Shutts, Banaji, & Spelke (2010), in which children selected novel objects after having heard two characters of variable sex, age, and race state their own preferences. Children’s selections of these objects were recorded as a measure of their own preferences along social group boundaries. This work showed that 3-year-olds preferentially select items and activities endorsed by same-gender and same-age characters, showing that some types of group membership guide early social preferences with meaningful behavioral outcomes. Other work in a similar vein has shown that even infants use language and accent to guide their selection of objects, and two-year-old children use these same cues to decide whom to give gifts (Kinzler, Dupoux, & Spelke, 2012). Such work shows that from an early age, children may be quite willing to use perceptual cues to make inferences and form preferences about other people, even in the absense of any behavioral information that would guide such judgments.

Experiment 2 emulated the general method used in Shutts et al. (2010) with the aim of investigating whether children’s selection of novel objects is influenced by the facial features of the characters that appear (who are uniform in major social category membership, as all are White males of a similar age). The investigations that follow in Experiments 3, 4, and 5 revise this approach; rather than asking children whose preferences they would emulate, they instead ask children to choose which of two faces
they would prefer to give a small “gift.” Experiment 5 goes one step further by asking participants to both select the faces they think are “nice” and to choose to whom they would like to give gifts, allowing for a direct comparison between responses from the two tasks. By asking children to give objects rather than select some for themselves, Experiments 3-5 bear particular resemblance to Experiment 2 of Kinzler et al. (2012), in which 2-year-olds were asked to choose recipients for various objects based on the languages in which they spoke.

**Experiment 2**

*Method*

Experiment 2 carried out an initial investigation into face-based behavioral responses using an adapted version of the paradigm used in Shutts et al. (2010) to serve as our measure of social preference. Children viewed pairs of computer-generated faces varying in either trustworthiness or dominance, with each face pair consisting of either a trustworthy and an untrustworthy face or a dominant and a submissive face. These face pairs were chosen based on the robust “nice/mean” judgments given to faces varying in trustworthiness and dominance observed in Cogsdill et al. (2014). Face pairs alternated between “trustworthiness” and “dominance” trials, and the orders of face pairs for both types of trials were randomized. Critically, the “nicer” face appeared on the left and right sides with equal frequency (four trials each). Although the order in which item pairs were presented was fixed, their relative position on the left or right side was counterbalanced, yielding two different trial order conditions. The effect of this condition on overall responding was not significant ($t(38) = 1.27, d = 0.41, n.s.$).
Children viewed face pairs on sheets of paper tucked into sheet protectors within a three-ring binder. For each trial, the experimenter showed the child the two faces, told them which items the faces preferred, and then asked them which item they would rather play with (or eat, wear, and so on, depending on the type of item). The items that faces endorsed were represented using images identical to those used in Shutts et al. (2010; Figure 3.6), and were referred to by nonsense words like “spoodle” and “blicket” (Table 3.2). Children then selected a laminated token depicting the object of their choice. It was hypothesized that children would select items endorsed by the trustworthy or submissive faces in each pair (i.e., those that were previously selected as “nice” in Paper 1) more frequently than those endorsed by untrustworthy or dominant (i.e., “mean”) faces.

Figure 3.6. All pairs of novel object stimuli used in Experiment 2.
<table>
<thead>
<tr>
<th>Word 1</th>
<th>Word 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spoodle</td>
<td>Blicket</td>
</tr>
<tr>
<td>Merp</td>
<td>Garn</td>
</tr>
<tr>
<td>Babber</td>
<td>Kazoop</td>
</tr>
<tr>
<td>Dak</td>
<td>Fep</td>
</tr>
<tr>
<td>Borg</td>
<td>Hib</td>
</tr>
<tr>
<td>Trop</td>
<td>Rizz</td>
</tr>
<tr>
<td>Spoot</td>
<td>Fres</td>
</tr>
</tbody>
</table>

Table 3.2. Nonsense word pairs used to describe novel objects in Experiment 2.

Results & Discussion

The results from Experiment 2 generally supported this hypothesis. On average, the children who participated in this study ($N = 40$, all between the ages of 3 and 10) selected the item associated with the trustworthy or submissive face in 58% of all trials. This was significantly greater than chance ($t(39) = 2.35$, $d = 0.37$, $p < .05$). When trials using trustworthy/untrustworthy and dominant/submissive faces were analyzed separately, they yielded levels of consensus at 56% and 60%, respectively, (Figure 3.7), a difference that was not significant (repeated measures ANOVA: $F(1, 39) = 0.52$, $\eta^2 = 0.013$, n.s.). In addition, patterns of response were consistent across development. When regressing consensus on children’s age in years, a statistically non-significant relationship emerged, $R^2 < .001$, $B < .001$, n.s.

Overall, while these data were consistent with the original hypothesis that children would be more likely to select the item endorsed by the trustworthy or
submissive face, they were also weaker than we had anticipated based on results from Shutts et al. (2010), which reported upwards of 70% of 3-year-olds choosing items endorsed by same-race characters who also matched the participants’ gender and age group (child vs. adult). Two features of the adaptation of the Shutts et al. procedure that was used here might have accounted for the lower-than-expected consensus that was observed. First, children might have been ignoring the faces entirely, and choosing instead to simply select the items they preferred based on their visual appearance. Although efforts were made to closely match the procedure used by Shutts et al. (2010), it is possible that some of the differences in the procedure used here had the unintended consequence of causing children to direct less attention to faces than was achieved in prior research. For example, faces in prior work “spoke” directly to children through recordings of human voice actors; here, however, the experimenter explained to children that the different faces preferred different items, rather than the characters speaking for themselves. In addition, due to the way in which items were set up for children to select, some children might have interpreted the behavior not as choosing an item for themselves, but of actually taking the item away from one of the faces. If children had interpreted the task in this way, they might have “taken away” items from faces that they dispreferred, thereby causing them to respond in patterns directly opposing those that we had hypothesized.
Figure 3.7. Overall findings obtained in Experiment 2 for all faces, followed by results for dominance and trustworthiness trials when analyzed separately. Symbols indicate two-tailed significance of difference from random responding (50%): * $p < .05$.

Experiment 3

Method

In light of the limitations we identified in the procedure used in Experiment 2, a different approach to study children’s face-based social preferences was implemented in Experiment 3. As in Experiment 2, children in Experiment 3 made decisions about pairs of faces presented in a three-ring binder. However, rather than choosing an item endorsed by one of the faces, children were given a laminated picture of a desirable item, such as a cookie or a chocolate bar (see Figure 3.8 for the full set of stimuli), and asked to indicate to whom they would rather give the item. In each trial, the experimenter used the following script, with different names appearing in each trial: “This is Edgar, and this is Martin. If you had only one cookie, which person would you give it to? Edgar, or

---

10 Images were obtained from a Google image search of clipart in the public domain.
Circular pieces of Velcro beneath the face images and on the backs of the tokens enabled children to respond by sticking the token directly onto the space beneath the face of their choice. Two practice trials using dogs and cats as targets, with steak and a toy mouse as sample “gifts,” ensured that children understood the procedure prior to beginning the experiment. The main experiment consisted of a total of eight trials, with four each displaying untrustworthy/trustworthy and dominant/submissive face pairs. Each individual face appeared only once throughout the experiment. All participants gave faces a total of eight different gifts, which were presented in a fixed, predetermined order.

In Experiment 3 alone, various hairstyles were added to face images using Adobe Photoshop. This was an exploratory technique that was intended to make stimuli more variable and lifelike to children, both to make the task more interesting and to elicit more naturalistic judgments. Some examples of these altered face images are provided in Figure 3.9.

Figure 3.8. Gift stimuli used in Experiments 3-5. (Experiment 5 used only the first four items: apple, banana, candy, and cookie).
Results & Discussion

A total of 89 children between the ages of 3.06 and 10.67 (Mage = 6.24, SD = 2.06) participated in Experiment 2, with 48 females, 37 males, and 4 whose parents declined to specify their gender. Overall, participants gave items to the trustworthy or the submissive face (i.e., the “nicer”-looking one) in 64% of all trials, which was significantly greater than chance ($t(89) = 5.83, d = 0.64, p < .001$). When analyzed separately, trustworthiness and dominance trials showed nearly identical levels of consensus at 63.9% and 64.4%, respectively, and this minor difference was not statistically significant ($F(1,89) = 0.03, n.s.;$ see Figure 3.10). As in Experiment 2, age in years did not significantly impact the percentage of trials in which children gave items to the “nicer” face, $R^2 = .01, B = .01, n.s.$
Figure 3.10. Overall findings obtained in Experiment 3 for all faces, dominance trials, and trustworthiness trials.

The convergence in findings between Experiments 2 and 3 was encouraging, and suggested that our preliminary findings, though relatively weak in magnitude, likely reflect a meaningful and replicable phenomenon. Taken together, both sets of findings lend robust support to the idea that children not only attribute traits and behaviors to faces, but that these inferences might also relate to social behavioral responses that children generate. Experiment 3 in particular suggests that children’s prosocial behaviors towards others might be influenced by the first impressions they form based on physical appearance. This finding complements the growing body of work showing that even 18-month-old infants produce spontaneous prosocial behaviors (Warneken & Tomasello, 2006) by suggesting that such behaviors might be modulated by children’s first
impressions of others, even when those impressions are based on as impoverished a stimulus as static pictures of faces.

Although the overall level of consensus observed in Experiment 3 was significantly greater than chance, it was not significantly greater than that observed in Experiment 2, $t(128) = 1.40, d = 0.23$, n.s.. Additionally, some children spontaneously verbalized rationales behind their choices that suggested they were following certain uncommon strategies. For example, one child verbalized their intent to give each gift to the face that appeared sadder. It is also possible that children might have deliberately given gifts to more threatening looking faces (i.e., untrustworthy and dominant) as a hypothetical gesture of appeasement. Thus, it is possible that some children’s choices did not reflect their social preferences per se, and may have even reflected the faces they preferred less.

**Experiment 4**

**Method**

Following the consistent results that emerged in Experiments 2 and 3 despite a number of possible sources of inconsistency in children’s responses, Experiment 4 carried out a conceptual self-replication after making several improvements to the experimental protocol. First, whereas face stimuli in Experiment 3 appeared in the same pairings across the four different stimulus orders that were used, in Experiment 4 the faces were paired differently across the four different order conditions that were used\(^{11}\). In addition, faces in Experiment 3 had hair added to them using Adobe Photoshop. This was an exploratory

---

\(^{11}\) Because the face sets included a total of 16 faces, with four of each type of face (trustworthy, untrustworthy, dominant, and submissive), a total of 16 order conditions would be necessary to include all possible two-way combinations of faces varying in dominance and trustworthiness. For the sake of practicality, we propose to continue using only four different face order conditions; however, if significant differences emerge between these conditions, it will be necessary to add additional conditions to minimize the effects of variance induced by unexpected results generated from certain face combinations.
technique that was intended to make the computer-generated faces appear more lifelike, as all faces were originally generated without any hair by default. However, in Experiment 4, these faces were replaced with the original (bald) versions of faces in order to eliminate variance in judgments that might have been caused by different hairstyles. The face identities were the same as those used in Experiment 3, and both the set of gifts to give to the faces and the order in which gifts were presented were also held constant.

**Results & Discussion**

A total of $N = 100$ children between the ages of 3 and 13 ($M_{age} = 7.15$, $SD = 2.37$) participated in Experiment 4. The sample was predominantly White (83%) and female, with 72 females and 28 males. Across all participants, children gave their gifts to the trustworthy or submissive faces (i.e., the “nicer” ones based on prior data) in 68% of all trials, which was significantly greater than chance responding, $t(99) = 8.11$, $d = 0.82$, $p < .001$. In a manner consistent with results from Experiment 3, this number was not significantly different between trustworthiness trials (69%) and dominance trials (68%, $t(99) = 0.33$, $d = 0.07$, n.s.), and responses to both types of trials were significantly more consistent than chance ($ts > 6.66$, $ds > 0.66$, $ps < .001$; see Figure 3.11).
**Figure 3.11.** Overall findings obtained in Experiment 4 for all faces, dominance trials, and trustworthiness trials.

The effects of age on consistency when analyzed using Pearson correlations were also consistent with the results obtained from Experiment 3. Across all trials, the effect of age on the percentage of trials in which gifts were given to trustworthy or submissive faces was statistically significant, $r = 0.26, p < .01$. Significant effects of age were also observed in dominant/submissive and trustworthy/untrustworthy trials when analyzed separately ($r_s > 0.20, ps < .05$).

Additional analyses explored whether children in the younger half of the sample responded with significant consensus. A median split was performed on the data, which were analyzed in two groups consisting of the youngest (ages 3-6, $n = 48$) and oldest participants (ages 7-12, $n = 52$). This analysis revealed that overall consensus remained
highly robust for both younger children (65% consensus; \(t(47) = 4.41, d = 0.64, p < .001\)) and older children (71%; \(t(51) = 9.71, d = 1.00, p < .001\)).

These results suggest that the behavioral propensity to preferentially give gifts to “nicer”-looking faces, whether they appear more trustworthy or more submissive in appearance, is one that both emerges early in childhood and also becomes more robust with age. The convergence in findings between verbal tasks used in Experiment 1 as well as prior studies (e.g., Cogsdill et al., 2014) and the behavioral tasks used here suggest that children’s face-to-trait inference goes beyond merely answering questions about extreme faces, and may in fact translate to behavioral responses that are differentiated on facial appearance. Children’s face-to-trait inferences might actually influence their social interactions from as early as the preschool years, a fact that may have cascading effects throughout the lifespan.

**Experiment 5**

This final experiment was conducted to directly test whether children’s “nice/mean” judgments of faces predict their gift-giving behaviors towards those faces. One limitation of Experiments 3 and 4 was that they were unable to account for the possibility that a child might have given a gift to a “mean”-looking face because they actually perceive that face as looking nicer. In other words, responses that may have been interpreted as reflecting randomness in children’s responses may have in some cases been completely consistent with respect to their individual judgments of faces. Experiment 5 addressed this possibility using a within-subjects design in which all participants formed both
“nice/mean” evaluations of faces as well as giving gifts to some of them, using the same procedures that were outlined in the foregoing studies.

**Method**

All participants completed both a “nice” face selection task, which closely resembled that used in Cogsdill & Banaji (under review), as well as a gift-giving task that followed the same procedure used in Experiments 3 and 4. Children were told at the beginning of the experiment that they would be playing a game consisting of two “rounds,” with the first round of the game taking place on the computer and the second round in a nearby binder.

As in Experiment 4, face stimuli consisted of computer-generated faces designed to appear trustworthy (i.e., “nice”) or untrustworthy (“mean”), with no additional hair added to the face images. In order to accommodate two rounds of experimentation in the short amount of time available while testing each child participant at the Boston Children’s Museum, only faces varying in trustworthiness (and not dominance) were used. Round 1 (“nice” judgments) consisted of eight trials in which eight trustworthy and eight untrustworthy faces appeared on a computer screen, in a similar fashion as in Experiment 1. Round 2 consisted of only four trials in which four each of trustworthy and untrustworthy faces appeared in pairs in a binder, in the same fashion as Experiments 3 and 4. All eight of the faces that appeared in Round 2 were also used in Round 1. The remaining eight faces in Round 1 that did not appear in Round 2 were included to obscure the repetition of faces across the two rounds. Although it is possible that some children may have noticed this, no participants mentioned the repetition of faces between rounds.

In Round 1, children saw pairs of faces presented on the computer screen and were instructed to point to the face in each pair that they thought looked “nice.” As in
Experiment 1, these experimental trials were preceded by three “big/small” practice trials that ensured that children understood and were attending to the task. After children completed all eight trials of Round 1, they were told that it was time to begin Round 2. The experimenter set the laptop aside and retrieved the binder containing stimuli for Experiment 5, which were identical to those of Experiment 4 but with all dominance trials removed. In addition, only the first four of the eight “gifts” used in Experiment 4 were used in Experiment 5 (apple, banana, cookie, and cupcake; see Figure 3.8). All children who began the experiment completed both rounds and their responses were included in the analyses that follow.

**Results**

A total of $N = 43$ children participated in Experiment 5, with 9 males, 33 females, and 1 whose gender was unspecified. Children ranged in age from 4 to 11.75 years of age ($M_{age} = 7.25, SD = 2.24$) and the majority was White (60%).

Overall results from both Rounds 1 and 2 are summarized in Figure 3.12. In Round 1, children demonstrated a strong tendency to select the trustworthy face as the “nice” one, doing so in an average of 85% of trials ($SD = .20$), which was much greater than chance ($t(42) = 11.35, p < .001, d = 1.75$). This rate of consensus did not correlate significantly with age ($r = .03, n.s.$) and thus likely reflects the upper boundary of children’s ability to respond consistently. In Round 2, children gave their gifts to the trustworthy faces in 72% of all trials ($SD = 0.34$), which was also greater than chance ($t(42) = 4.30, p < .001, d = 0.65$), though this level of consensus was significantly lower than that obtained in Round 1 ($t(41) = 2.94, p < .01, d = 0.92$). The effect of age on
consistency in Round 2 was larger than in Round 1, but still did not reach statistical significance \( (r = 0.26, n.s.) \).

Figure 3.12. Overall findings obtained in Experiment 5 in Round 1, in which children selected the face in each pair that was “nice,” and in Round 2, in which children decided which faces to give small “gifts.”

The within-subjects design of Experiment 5 enabled us to analyze the correspondence of judgments and behaviors between the two rounds. This was measured as the percentage of faces (among the sixteen faces that appeared in both rounds) for which each participant made the same selection\(^\text{12}\). Thus, even if participants had atypical opinions about who appeared nice or mean, the degree to which these judgments influenced their giving behaviors in Round 2 would determine whether this number was

\(^{12}\) Note that although these 16 faces appeared in both rounds, they were nearly always paired differently between them, as pairings were determined randomly in Round 1 without deliberately matching (or avoiding) the pairings used in Round 2.
high or low, with a number of 50% indicating a lack of any systematic tendency to be
generous to the faces that were selected as nice. On average, children gave to the faces
they judged to be “nice,” and declined to give to the faces they did not select as nice, for
76% of all faces (8 faces per child). This was greater than chance (50%; \( t(42) = 6.21, p < .001, d = 0.95 \)) and did not change significantly with age (\( r = 0.27, n.s. \)).

**Discussion**

The results from Experiment 5 provide yet more corroborating evidence to suggest that
children modulate their behaviors based on the facial appearance of others and, it would
appear to follow, their face-based social judgments of others as well. These results
replicate those obtained in Experiments 3 and 4, a fact that is even more impressive given
the smaller sample size used in this final experiment (\( N = 43 \) versus \( N = 89 \) and \( N = 100 \)
children in Experiments 3 and 4, respectively; see Figure 3.13 for a comparison across
the three experiments). The self-replication in Experiment 5 is powerful given its use of
only four trials in the gift-giving task, compared to the eight trials used in prior
experiments. Although the use of half as many trials and participants might have been
expected to generate substantial noise in the data, the findings that emerged clearly
corroborated those that had been obtained previously. This remarkable consistency
allows us to conclude with confidence that children show a consistent and measurable
tendency to behave prosocially towards faces, even ones that they view as computer-
generated, static, two-dimensional images.
Figure 3.13. Rates of giving gifts to nicer-looking faces across Experiment 3 (faces with hair), Experiment 4 (bald faces), and Experiment 5 (bald faces with “nice/mean” judgment task). Significance is displayed relative to chance responding at 50%, all $p < .001$.

The fact that such strongly convergent results were obtained from such stimuli might further suggest the possibility that behaviors in real-world contexts, in which social targets are real people, may be affected even more strongly by children’s judgments based on others’ facial appearance. Further studies using increasingly rich stimuli – perhaps even including live actors as targets – as well as meaningful, ecologically valid measures of children’s prosociality will be necessary to explore this possibility. Such work will be important in determining the full extent to which early face-based social judgments affect children’s interactions with others, which bears the potential for deep and pervasive effects throughout a child’s early social development.

By asking children to produce behaviors rather than verbalize judgments, the methods used in Experiments 2-5 possess a significant advantage in that the responses
children generate are relatively independent of their verbal development and resulting interpretations of our questions, which can vary dramatically among the ages that were tested. Although all children must be verbally competent enough to understand the instructions of the task in order to participate, the act of selecting items or giving gifts to different recipients does not depend on individual interpretations of words such as “nice” or “strong,” an aspect of prior research designs that might have artificially reduced the consensus that was observed. This shift towards methods that are increasingly nonverbal in nature will be important in future research for clarifying our understanding of how children respond to faces well in advance of their ability to clearly articulate their impressions. Such work would complement the growing body of literature showing that even 18-month-old infants produce spontaneous prosocial behaviors (Warneken & Tomasello, 2006) by suggesting that such behaviors might be modulated by children’s first impressions of others, even when those impressions are based on stimuli as impoverished as static face images.

**General Discussion**

The experiments reported throughout this paper clearly demonstrate that young children form consistent judgments about others based on facial appearance, thereby providing strong corroborating evidence supporting prior work on this topic. Crucially, however, this collection of findings adds the additional insight that such attributions also extend into the behavioral realm in two ways: first, in the form of attributions of behavior, and second, in the domain of behaviors that children actual perform in social contexts. This combination of results serves to highlight the importance and real-world relevance of
prior work, which hitherto only demonstrated the precocity and generalizability of face-based trait inferences. In essence, while such work had shown that children are capable of making adult-like trait inferences about faces, the present studies suggest that these inferences extend to behavioral attributions and may even affect how children decide to behave towards other people.

Further research will be necessary to clarify the full extent of these effects using novel experimental designs, particularly those with a focus on improving external validity by more closely imitating real-world social interactions. An important advance in recent research has been the inclusion of behaviors from social agents – for example, in Ewing et al. (2014), in each trial the (fictional) recipient gives back some amount of money that the child bestows on them. Interactive methods like this may help the social experiences that occur within the context of psychological experiments to feel more meaningful and relevant to children, thereby eliciting behaviors that are more representative of everyday interactions. Similar advances in experimental design in the future, perhaps even involving the use of more dynamic stimuli, such as videos or even live actors, will allow for increasingly meaningful insights about the impact of facial appearance on children’s social judgments and behaviors early in life.

Face-to-trait inference is not an isolated phenomenon. Rather, as the present work demonstrates, children’s early face-based social judgments go beyond mere trait inferences. They guide children’s expectations about how others behave, and they influence their decisions about how to behave towards others. While decades of research has shown that adults use appearance-based judgments to guide many important social decisions, the present work provides clear evidence that the consequential nature of such
judgments extends as far back as the preschool years. These results suggest a need for new methods to be developed for studying progressively younger populations, and using more ecologically valid tasks, for the ultimate goal of understanding how social judgments initially emerge and work to shape social development throughout the lifespan.
CONCLUSION
The predisposition to make rapid and unreflective judgments based on minimal facial information is a pervasive form of social judgment. Past research focusing on adult judgments has suggested that such inferences have important and ubiquitous real-world consequences, and the pervasive effects of such judgments motivates the present study of their emergence and development early in childhood. Prior to the work included in this dissertation, little was known about whether such inferences were unique to adults, requiring prolonged social and perceptual experience to manifest, or whether they instead reflect deep-seated intuitions guided by a fundamental and early-emerging element of human social cognition.

The three papers that comprise this dissertation provide compelling and consistent evidence that face-to-trait inference is a fundamental social cognitive capacity that develops in a robust manner and even begins to influence social behaviors early in childhood. The first paper, Cogsdill et al. (2014), showed that children as young as 3-4 years of age show a strikingly adult-like tendency to categorize faces along dimensions of “nice/mean,” “strong/not strong” and “smart/not smart,” indicating that such inferences are deeply rooted early in social cognitive development. The second, Cogsdill & Banaji (under review), demonstrated that this early form of face-to-trait inference is highly robust, with cross-age consensus emerging strongly in response to naturalistic stimuli that include faces belonging to adults, children, and even rhesus macaque monkeys. The third and final paper, Cogsdill, Spelke, & Banaji (in preparation), highlights the significance of these findings by showing that early-emerging social judgments based on faces influence both children’s expectations of how others are likely to behave as well as their own
behavioral tendencies towards others who possess certain configurations of facial features.

Taken together, these papers paint a clear picture: face-to-trait inference develops early and influences other social judgments and even behaviors early in childhood. The set of findings presented here suggest that face-to-trait inferences take root early in development and exert a powerful influence on social judgments, interactions, decisions, and behaviors, with effects that cascade throughout the lifespan and consequences that pervade society.

Because these studies represent some of the earliest work exploring the development of face-to-trait inferences early in childhood, there exist many possible avenues for future research to take to more thoroughly understand this important social cognitive capacity. One frontier that is ripe for exploration is that of studying whether infants also demonstrate signs of adult-like intuitions about the relationship between facial features and personality traits. The studies in this dissertation study children as young as three years old, which is the earliest age at which they can reliably complete tasks that require them to understand and follow basic instructions and offer simple verbal responses. However, future work that studies even younger populations, particularly babies, will be critical to forming a more complete understanding of when stereotypes concerning facial structures begin to cement in the mind of the developing child. If such work were to successfully overcome the challenges that come with studying infants using purely nonverbal measures such as preferential looking time – for example, the need to control for preferences driven by factors like facial attractiveness and femininity – it would produce insights of crucial importance to our understanding of the
earliest development of face-to-trait inferences. While the skills that infants possess in merely perceiving faces are well-documented, the existence of primitive character judgments in the earliest stages of social cognitive development would represent a significant advance in our understanding of the primacy with which such judgments develop early in life.

Other areas of future study that will greatly benefit this emerging body of research will be those that continue to broaden the set of stimuli and tasks used to probe early face-based person impressions. One limitation of the work presented in this dissertation is that all studies use exclusively static, two-dimensional stimuli. In the spirit of Cogsdill & Banaji (2014), which demonstrated the broad generalizability of early face-to-trait inference by using diverse sets of face stimuli belonging to entirely different groups of people and even non-human primates, future work will benefit greatly from incorporating an even wider array of stimuli and tasks. For example, such work might find a way to incorporate dynamic stimuli in which either computer-generated or naturalistic faces display some kind of motion, thereby taking an approach that more closely resembles real-world interactions involving living and moving interaction partners. By striving to make innovative experiments that truthfully emulate real-world interactions, future researchers can gain an even clearer understanding of the forces at play in the mind of society’s youngest social agents.

The study of early appearance-based social judgments may also be extended beyond the realm of the human face. For example, following from work that has explored children’s use of bodily cues when interpreting ambiguous displays of facial emotion (Mondloch, 2012), a similar approach could conceivably be adapted to study whether and
to what degree children use such cues to render even simple judgments as whether to approach or avoid an unfamiliar individual. By expanding the topic of inquiry to include aspects of appearance other than the face, researchers can achieve a more comprehensive understanding of the nature of early-emerging social judgments. Such work will also reveal whether the face is a particularly strong signal of personality traits early in social development, or whether the tendency to connect physical appearance to character traits is so essential to early social cognition that it is readily applied to any salient aspect of physical appearance.

Future work might also explore the interaction between structural features and known stereotypes, such as those relating to sex or race, following the lead of recent research that has explored the nature of facial dominance in females (e.g., Sutherland, Young, Mootz, & Oldmeadow, 2014). Such an approach would serve to further clarify the precise role of facial physiognomy in guiding the social judgments of both children and adults across the diverse social interactions that perceivers encounter in everyday life.

The experiments reported throughout this dissertation follow a long tradition of research that has consistently demonstrated the powerful effects of first impressions. From the moment a person first encounters another, they form countless perceptions and judgments about them, all of which contribute to decisions that, once accumulated over the billions of people that have ever lived, combine to create the fabric of the world we now inhabit. This dissertation contributes to this existing field the consistent and robust finding that face-to-trait inference, one of the cognitive propensities that underlie such
social judgments, is so fundamental to human social cognition that it reaches a consistent and mature state early in childhood.

Still remaining is the question of how and why this propensity to use facial features to evaluate character traits originally came to be a fundamental component of human social cognition. While this dissertation has primarily concerned itself with the question of ontogeny – that is, when in the course of human development face-based character judgments emerge – it is also important to consider from a phylogenetic perspective why these judgments are so deeply rooted in the human psyche. One explanation that has been offered for the existence of face-to-trait inferences is the overgeneralization hypothesis, which stems from an ecological theory of social perception. Under this theory, social impressions served a predominantly adaptive function over the course of evolutionary history by enabling perceivers to make accurate inferences of social characteristics like health and emotional state. However, the same features that guide these accurate perceptions are readily overgeneralized to produce inferences that may not be accurate, such as evaluations of character traits.

Perhaps the clearest example of this is in perceptions of trustworthiness, an enduring character trait that is overgeneralized from features that resemble expressions of happiness, a temporary, approach-oriented emotional state. Both behavioral and neuroimaging data (e.g., Todorov, 2008) support the idea that perceptions of trustworthiness based on facial appearance constitute an overgeneralization of features signaling happiness (more trustworthy) or anger (less trustworthy). Thus, to the extent that emotional perception is itself a highly adaptive tool for rapid and nonverbal social communication, allowing for its proliferation within our species, it may be that face-to-
trait inference is a side effect of an otherwise adaptive system for social perception, which may cause perceivers to render a number of inaccurate (but not fatal) judgments on a routine basis.

Some have argued that, besides simply being a side effect of an adaptive mechanism for social perception, facial features may themselves serve as honest signals to personality. Although the work in this dissertation deliberately set aside questions regarding the accuracy of face-to-trait judgments, this issue is of particular relevance when considering why such judgments may have originated as part of the battery of social judgments humans render in everyday life. Some have argued for the existence of a biological connection between certain forms of appearance and personality. Such a relationship might occur through the work of hormones like testosterone (for example), whose effects over the course of development may shape both facial features, such as facial width-to-height ratio, as well as behaviors that have been related to those features, such as aggression (e.g., Carré, McCormick, & Mondloch, 2009). Related to this, other work has reported significant associations between facial averageness and symmetry, both of which relate to overall attractiveness, and outcomes like composite health measures and intelligence (Zebrowitz & Rhodes, 2004). Such work that has been taken to support the “bad genes” and “anomalous face overgeneralization” accounts of face-based social judgments, which state that such judgments arose from an adaptive ability to accurately detect characteristics of adaptive significance, such as health and competence, using only a small number of visual cues available to perceivers.

Thus, if some facial features may serve as valid indicators of personality traits of survival value to others, such as dominance or aggression, one could claim that face-to-
trait inference represents an adaptive mechanism of social perception. If true, even if many inferences are inaccurate, to the extent that some have conferred adaptive benefits by allowing individuals to avoid threatening individuals, such judgments may have persisted in our species. This would be consistent with the early emergence of highly polarized trait judgments early in life, as not only mothers but also their young children would stand to benefit from detecting threats in the environment, even to the point of doing so erroneously to the extent that such judgments represent stereotypes that may be inaccurate.

In addition to biological factors, the mere existence of stereotypes may bring about a kernel of truth to such judgments through self-fulfilling prophecy effects. The classic finding by Rosenthal and Jacobson (1968) that teacher’s expectations of students’ competence predicted the performance of those students has inspired decades of work showing that mere expectations of others can profoundly impact how they actually behave. Self-fulfilling prophecy effects have been studied extensively as part of a larger research enterprise of investigating how expectations can serve to construct social reality (e.g., Snyder & Stukas, 1999; Willard et al., 2012). Whether face-to-trait stereotypes exist for biologically plausible reasons or are simply biases that cannot be traced to any potential sources of accuracy, it is possible that the mere existence of such stereotypes may bring about measurable and consistent relationships between facial features and personality traits, although the existence of self-defeating prophecy effects (e.g., Collins & Zebrowitz 1995; Zebrowitz et al. 1998) may complicate the path that leads from stereotypes to reality.
A final question concerning why face-to-trait inferences develop so early in life concerns that of where young, even pre-verbal children obtain the information necessary to develop these stereotypical associations between facial features and traits. Work with adults has shown that the human mind possesses ability to quickly generate implicit associations based on statistical regularities among even in novel and unfamiliar contexts (e.g., Gregg, Siebt, & Banaji, 2006). This, combined with data showing that even young children demonstrate implicit bias (e.g., Cvencek, Meltzoff, & Greenwald, 2011; Dunham, Baron, & Banaji, 2006; Dunham, Chen, & Banaji, 2013), suggests that even a young child with only a couple of years of social experiences might have had sufficient opportunities for associative learning to generate facial stereotypes early in development. Even a three-year-old has had many rich experiences both directly through interactions with family members and other adults and children, as well as indirectly through exposure to media. One possibility, while speculative in nature, is that the stereotypical appearances of “good” and “bad” characters that children experience from the earliest years of life may contribute to the development of early face-to-trait associations. Future research will be necessary to understand what sources of social information provide the original basis for learning face-based stereotypes early in childhood, and perhaps even in infancy.

Given the early-emerging nature of face-to-trait inferences, they are likely to influence social judgments and interactions throughout childhood, with effects that cascade throughout the lifespan. Children interact with family and other caretakers and educators from an early age, and these early experiences may be shaped by facial stereotypes in numerous and profound ways. These young social perceivers grow up into
adults, whose decisions based on their assessments of others have the power to determine relationships of all kinds – who gets hired or fired, admitted or rejected, convicted or acquitted, and so on in countless other examples throughout society. By undertaking a set of rigorous and cohesive explorations into the developmental origins and social consequences of face-to-trait inference, this dissertation provides a step towards understanding the ontogenetic origins of face-based character judgments, thereby beginning to fill a long-standing gap in the field of social cognitive development. These data show clearly that children show a robust and adult-like willingness to use superficial facial cues to render judgments of character traits, a finding that demands that future research continue to strive towards a more complete understanding of this fundamental social cognitive propensity.
REFERENCES


doi:10.1111/j.1467-9280.2009.02423.x


doi:10.1080/02699931.2012.724012


doi:10.1177/0956797614523297

doi:10.1371/journal.pone.0069684


doi:10.1207/S15327957PSPR0402_01


doi:10.1177/0956797610362647


doi:10.1111/bjop.12085


doi:10.1126/science.1199198.


Emotion, 13(4), 724–738. doi:10.1037/a0032335


