Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Sara Valencia Botto Date

Emergence and Cognitive Correlates of Evaluative Audience Perception

By

Sara Valencia Botto
Doctor of Philosophy

Psychology

_____

Philippe Rochat, Ph.D.
Advisor

_____

Frans de Waal, Ph.D.
Committee Member

_____

Robyn Fivush, Ph.D.
Committee Member

_____

Stella Lourenco, Ph.D.
Committee Member

_____

Elaine Walker, Ph.D.
Committee Member

Accepted:

_____

Lisa A. Tedesco, Ph.D.

Dean of the James T. Laney School of Graduate Studies

_____

Emergence and Cognitive Correlates of Evaluative Audience Perception

By

Sara Valencia Botto
B.S., University of Georgia, 2014
M.A., Emory University, 2016

Advisor: Philippe Rochat, PhD

An abstract of
A dissertation submitted to the Faculty of
the James T. Laney School of Graduate Studies of Emory University
In partial fulfillment of the requirements for the degree
of Doctor of Philosophy in Psychology
2020

Abstract

Emergence and Cognitive Correlates of Evaluative Audience Perception

By Sara Valencia Botto

For the past few decades, research has documented how reputational concerns change adults' and children's behavior across contexts. For instance, in public, but not in private, adults and children are significantly more generous, conform to the opinion of the majority, and make explicit attempts to manage their reputation. Clearly, the concern for others' potential evaluations influence much of human behavior. However, despite this concern being central to human psychology, little is known about its ontogeny. In particular, when do children begin to show first signs of a concern for the potential evaluation of others, what we refer to here as *evaluative audience perception*? And what does it take to develop such a concern? Using a developmental perspective, the present dissertation explored potential social-cognitive abilities — including children's objectified sense of self, their theory of mind, and their norm understanding — that might predict the emergence of evaluative audience perception in 14- to 27-month-old children (N = 59). Specifically, we hypothesized that if these three social-cognitive abilities are potential pre-requisites of evaluative audience perception, then children who showed all three social-cognitive abilities would be more likely to strategically modify their behavior based on the previously demonstrated values of the experimenter, as well as her relative attention (looking versus not looking at the child). In support of this hypothesis, results showed that only children who displayed theory of mind, an objectified sense of self, and normativity, strategically modified their behavior depending on whether they were being observed. In particular, we found that — independent of age, general cognitive competency, or temperament — children who demonstrated all three social-cognitive abilities were significantly more likely to reproduce an outcome that had been positively valued by the experimenter when the experimenter was watching, but then chose to reproduce an outcome that had been *negatively* valued by the experimenter once the experimenter had turned her back. On the whole, this dissertation sheds light on a potential cognitive mechanism underlying the human concern for reputation.

Emergence and Cognitive Correlates of Evaluative Audience Perception

By

Sara Valencia Botto
B.S., University of Georgia, 2014
M.A., Emory University, 2016

Advisor: Philippe Rochat, PhD

A dissertation submitted to the Faculty of
the James T. Laney School of Graduate Studies of Emory University
In partial fulfillment of the requirements for the degree
of Doctor of Philosophy in Psychology
2020

Acknowledgements

The current dissertation would have not been possible without the invaluable guidance of my adviser and my committee members, the support of my laboratory members and family, and the participation of parents and children. I want to give special thanks to my adviser, Dr. Philippe Rochat, whose creativity, vast knowledge, and encouragement shaped me into the scientist I am today. I deeply appreciate the freedom you gave me to explore questions I was passionate about, and for the opportunities you afforded me. I also want to thank Dr. Stella Lourenco and Dr. Robyn Fivush, who beyond serving as my committee members, have been constant role models and mentors. It is your guidance and genuine care for me that challenged me and inspired me to achieve excellence, and I am forever indebted to you both. Thank you to Dr. Frans de Waal and Dr. Elaine Walker for your insight and valuable feedback during the development of this dissertation. It is an honor to have faculty of your caliber on my committee.

I also want to thank Dr. Daniel Dilks, who always offered me encouragement and pushed me to think critically about my study designs. To my Infant and Child Laboratory members, Natalie Eldred, Amber Wallace, Maria Jones, Shen Sheng Wang and Cynthia Guo, I am so thankful for the friendships we have developed and for the support you gave me throughout my graduate school career. You all are selfless and brilliant people, and I am looking forward to all you will accomplish. Thank you to all the research assistants, who spent countless hours recruiting participants and coding videos. I also want to give a special thanks to Vlad Ayzenberg, who has given me feedback on various projects, and the rest of my fellow psychology graduate students for making this journey enjoyable

Lastly, I want to thank my family, who gave me the strength to follow my dreams and accomplish my goals. To my siblings, Catalina and Andy, thank you for leading the way, and for always encouraging me and making me laugh when things were difficult. To Missy, Dennis, and Lauren, thank you for your constant support. To my amazing husband, Brett Botto, it is your unwavering love and support that got me through graduate school. You are my greatest blessing and I love you more than words could ever express. Finally, I would like to dedicate this dissertation to my parents, William and Luz Valencia, who selflessly immigrated to this country so that my siblings and I could get the best education possible. Mom and dad, you taught me what perseverance, dedication, and love look like. I would certainly not be here if it were not for your love, courage and prayers. I hope that I have made you proud.

Table of Contents

Emergence and Cognitive Correlates of Evaluative Audience Perception

Compared to other animals, humans have the peculiar propensity to adorn and present themselves in ways that will elicit positive evaluations from others. As case in point, the cosmetics industry, which exists to superficially enhance individuals' attributes while hiding their imperfections, is a multimillion-dollar industry (Termini & Tressler, 2008). Interestingly, the propensity toward body adornment can be traced back to the Neandertals (Zilhão et al., 2010). As a species, we display a unique concern with reputation (i.e., the computation of how others perceive and evaluate us; Rochat, 2013).

Given that our concern with reputation is central to human psychology, it is not surprising that reputational concerns are evident by 3-5 years of age (Engelmann & Rapp, 2018). For example, Zhao et al., (2018) showed that, when primed with a reputational cue (i.e., "I heard your peers think you're smart"), 3-5-year-old children are more likely to cheat in order to maintain a smart reputation. Numerous studies have also shown that, like adults, preschoolers are significantly more generous in public compared to in private (Engelmann et al., 2012; Piazza, Bering & Ingram, 2011), and will also conform to the opinion of the majority publicly but not privately (Haun & Tomasello, 2011). Clearly, reputational concerns affect much of human behavior.

However, despite the fact that reputational concerns are a critical part of human development, there is a paucity of work on its origins in human development (Silver & Shaw, 2018). Specifically, when do children begin to show first signs of a concern for the potential evaluation of others? And what does it take to develop such a concern? Using a developmental perspective, this dissertation aimed to explore possible social-cognitive abilities that might be related to the emergence and development of our concern with the potential evaluation of others, herein referred to as *evaluative audience perception*.

To provide some background, I first review literature on reputational concerns in both adults and children and argue that the way in which reputational concerns has been studied in developmental research eludes the question of its ontogenesis. Importantly, I argue how *evaluative audience perception* — a novel concept that provides a new framework on the developmental origins of reputational concerns — emerges by the second year and is foundational to the development of reputational concerns. Following this broader theoretical introduction, I then discuss open questions and motivations for the present study and discuss how three specific social-cognitive capacities — including an objectified sense of self, theory of mind, and normativity — could underlie the development of evaluative audience perception. Lastly, a behavioral study testing this hypothesis in fifty-nine 14-27-month-old toddlers is reported and results are discussed in the context of plausible developmental models and future directions.

Reputational Concern in Adults and Children

In his seminal 1959 book *The Presentation of Self in Everyday Life*, sociologist Ervin Goffman describes the various strategies humans adopt to manage one's image. Goffman uses the metaphor of a theatrical performance to describe individuals' propensity to "act" in social interactions to manage their persona (Goffman, 1959). Since then, psychologists have documented adults' reputation management through various reputational tactics, including ingratiation (Jones, 1964), conformity (Asch, 1956), and self-presentation (Baumeister, 1982). Though under different names, all of these constructs capture the significant effect of an audience on adult behavior. In public, compared to in private, adults tend to be more generous (Dana, Weber & Kuang, 2007), make riskier investments (Goulart, da Costa, Andrade & Santos, 2015), or conform to a majority opinion, even if this opinion is blatantly wrong (Asch, 1956).

These findings index how individuals strategically modify their behavior to manage their public image, and ultimately their reputation.

While there is ample evidence of reputation management in adults, it is not until recently that a handful of studies have explored when children begin to display reputational concerns. Converging evidence show that between the ages of 3 and 5, children's reputational tactics begin to look quite similar to those of adults. In a replication of Asch's (1956) classic study, Haun and Tomasello (2011) show that 4-year-olds, like adults, tend to conform to a majority opinion in public, but not in private. By five, children expect someone who is ostracized and seeking affiliation with a group to strategically engage in strong conformity (Cordonier, Nettles, & Rochat, 2018). Five-year-old children also tend to be more generous (Engelmann et al., 2012; Leimgruber, Shaw, Santos, & Olson, 2012), and cheat less in the presence of an observer (Engelmann et al., 2012). Various studies also suggest that by the end of the preschool years, children display sensitivity to reputational cues. For instance, when told that their peers thought positively about them, five-year-old children were less likely to cheat compared to when not primed with a reputational cue (Fu, Heyman, Qian, Guo & Lee, 2016).

Other studies indicate that even three-year-olds may already be sensitive to reputational cues, sharing more when primed with eyes (Kelsey, Grossman, & Vaish, 2018) and being more likely to cheat when told they have a reputation for being smart (Zhao et al., 2017). By the age of 6-8 years, children's attempts at managing their reputation becomes more explicit. They manifest modesty, flattery, or tell white lies to portray a certain image of themselves (Heyman, Fu, & Lee, 2008; Watling & Banerjee, 2007). In all, these studies suggest that from 3-4 years of age, children begin to be aware that their behavior is or might be evaluated by others. They start

making explicit efforts to manage their reputation and such efforts become increasingly sophisticated with development.

Evaluative Audience Perception as the Foundation for Reputational Concerns

Although current research captures when children begin to engage in reputational tactics, little is known about when and how reputational concerns might develop. Specifically, how does an infant — who seemingly does not care about the slobber on their face — develop into an individual who will spend resources to manage their image? One reason this question has remained largely unanswered is because researchers have used, as a benchmark, adult reputational tactics, including strong conformity and self-presentation, to investigate the development of reputational concern in children. This approach fails to address the ontogenetic development of reputational concerns, because adult reputational tactics entail higher-order cognitive processes, such as an advanced self-concept, a sophisticated understanding of social standards, and more importantly, an understanding that others' judgements might have lasting social and affiliation consequences (i.e., being rejected; Rochat, 2010; 2018).

To mitigate this issue and address the ontogenetic question of reputational concerns, it is important to reduce terms like reputation or self-presentation into their constitutive components. By analogy, and much like chemists who break down substances into elements to understand their fundamental properties, it is necessary for psychologists to find the essential components of reputation and self-presentation to capture how humans come to care about others' evaluation. Accordingly, we attempt to reduce these terms by proposing that the two components of evaluative audience perception (EAP) — which include our understanding that others can and will evaluate our behavior, and consequently, the concern for the valence (positive versus

negative) of this potential evaluation — are the necessary elements for the development of reputational concerns.

As an illustration, consider an individual engaging in self-presentation. To do this, the individual either through words, behavior, or self-adornment tries to convey a particular image to others. However, for this behavior to be worthwhile, the individual must first understand that others can or will evaluate such behavior. Otherwise, there would be no need to engage in reputational tactics. Likewise, if the individual did not care about the valence of the potential evaluation, either positive or negative, then there would be no need to spend resources in managing their image, with adornment or by displaying more generosity when others are watching. The same logic applies to any form of reputation management, from wearing make-up to telling white lies; all of these require both the basic ability to see others as evaluators and the tendency to care about the valence of the potential evaluation.

Following this rationale, we propose that the two components of EAP underlie reputational tactics, and therefore a child who engages in ingratiation, self-presentation, or shows reputational concern can do so *because* they have developed EAP. This proposal offers a different perspective on the study of the origins of reputation, as it shifts the focus from exploring when children begin to engage in adult reputational tactics, to exploring when and how infants begin to see others as evaluators of their own behavior. Such a perspective could elucidate the rudimentary prerequisites of reputation, and what allows for this human trait to emerge in development.

Beyond providing a different developmental lens for the development of reputational concern, EAP might also address the reason behind our general concern with others' evaluation. Some psychologists posit that humans care about reputation because it entails potential social

consequences (Leary & Allen, 2011; Tomasello, 2018). Indeed, someone who is considered generous is going to be more liked and accepted in a group than someone who is stingy. But what about instances in which there is no clear social consequences? For example, an individual might hesitate to take more than one free sample at the grocery store to not appear greedy. This scenario does not yield any direct consequences, yet the thought of being evaluated might drive such behavior. Accordingly, the development of EAP would serve as a basic heuristic guiding behavior across contexts.

Along the same vein, it is important to distinguish between displaying EAP and behaving strategically to avoid punishment[1]. While in some instances, the concern for others' positive or negative evaluations might be driven by the anticipation of potential positive or negative *social* consequences (such as gaining approval from others or being rejected), this concern cannot simply be driven by fear of punishment. Following our previous example, being more generous in public versus private is presumably not motivated by fear of being punished, but by a concern for how others might evaluate and respond to generosity (or lack thereof). In this case, being generous might actually be rewarded, although this consequence is not guaranteed. Unlike fear of punishment, EAP would capture the unique propensity to care about others' *potential* evaluation, even in the absence of immediate positive or negative consequences.

On the whole, evaluative audience perception provides a new framework to study the ontogenesis of reputational concern. Instead of focusing on adult reputational tactics, it strips the phenomenon to two rudimentary components, facilitating research with younger children. With this in mind, understanding when and how children begin to see others as potential evaluators of

[1] Following the literature, we define punishment as "the presentation of an aversive or the removal of a positive" stimulus to deter individuals from behaving in ways deemed unacceptable (Dadds & Salmon, 2003, p. 70).

the self would be the first step toward elucidating the foundation and development of reputational concerns in early childhood.

Emergence of Evaluative Audience Perception in the Second Year

Because we propose that EAP is the foundation for reputational concerns, EAP should presumably emerge prior to reputational concerns. In support of this rationale, recent studies suggest that already by 24 months, toddlers are sensitive to how others might evaluate (respond or react) their behavior. Botto & Rochat (2018) tested 14-24-month-old infants to explore when children would begin to strategically modify their behavior when others are watching. In a novel paradigm called the robot task, an experimenter showed 14-24-month-olds how to activate a toy robot by pressing a remote control and expressing either a positive (yay! Isn't that great?) or negative (oh! Oops, oh no!) value. After this initial toy demonstration, the experimenter invited the child to play with the remotes, and then either watched the child (attentive condition) or pretended to read a magazine (inattentive condition). Across four studies, we found that toddlers tended to modify their behavior depending on whether the adult demonstrator was watching, and whether the adult had positively or negatively valued the remote action. In particular, when the adult was watching, children tended to activate the remote associated with a positive value significantly more. In contrast, if the adult previously expressed a negative value, children tended to wait until the adult turned her back to activate that remote.

Note that this selective behavior occurred in the absence of a potential punishment, since children's button-pressing behavior was never followed by an aversive consequence or any reprimand. If children's behavior were motivated by a fear of punishment, then we would have expected children not to engage with the negative remote, regardless of the condition. However, this was not the case. This pattern of results also casts doubt on the possibility that children's

behavior might be simply driven by a generalized positive or negative association with either remote. Again, children not only chose to play with the positive remote or avoid the negative remote across conditions but seemed to consider the experimenter's reaction toward the remote *and* whether or not she was watching them.

Further supporting this notion is that, when we removed the differential values from the remotes in a follow-up control study (i.e., we just simply said 'oh wow' after pressing either of the remotes), children's button pressing behavior no longer differed across attentive versus inattentive conditions. That is, when the experimenter no longer expressed opposing opinions toward pressing either of the remotes, children no longer modified their button-pressing behavior depending on the attention of the experimenter. This provides strong support that children's consideration of the experimenter's expressed values in the previous studies drove strategic button-pressing behavior.

In addition to demonstrating EAP by strategically modifying their behavior in the presence of others, other social-cognitive phenomena emerging in the second year further demarcate children's increased concern with others' potential evaluations. For example, by 18-21-months, children begin to show embarrassment in situations that might elicit an evaluation, such as seeing a mark on their face upon seeing their reflection (Lewis, Sullivan, Stanger, & Weiss, 1989), or when publicly failing in a task (Stipek, Recchia, McClintic, & Lewis, 1992). Toddlers also show rudimentary signs of conformity when passing the mirror mark test, leaving a mark on their forehead, if surrounding others also have the same mark on their forehead (Rochat, Broesch & Jayne, 2012). Eighteen-month-olds will also selectively imitate an adult who shows positive as opposed to negative emotions while modeling an action (Repacholi &

Meltzoff, 2007). Furthermore, it is around this age that children begin to manifest prosocial behavior, particularly if they are primed with social affiliation (Over & Carpenter, 2009).

In all, these other-regarding developments by the child's second birthday reveal an increased concern with how others might respond to the child's behavior, supporting the idea that by 24 months, children do demonstrate the two defining elements of EAP: (a) the tendency to implicitly assume that one's behavior or appearance could be, or will be evaluated by others either positively or negatively; and (b) the default preference to elicit positive as opposed to negative evaluations (reactions or responses) from others. It is the combination of these two simple components that would serve as the bedrock for reputational concerns documented in the preschool years. An important follow-up questions is, what allows these two components to emerge?

Cognitive Correlates of Evaluative Audience Perception

As presented above, converging studies indicate that evaluative audience perception is evident by 24 months (Botto & Rochat, 2018; Over & Carpenter, 2009; Repacholi & Meltzoff, 2007). What remains to be explored is, what cognitive capacities are necessary for children to become sensitive to the evaluation of others? Exploring this question using a developmental perspective is particularly advantageous because it allows researchers to apply an 'inverse-engineering' approach to a particular phenomenon or cognitive capacity, like EAP. A developmental framework presupposes that an ability does not simply appear without previous cognitive, physical, or emotional milestones (Damon, 1983; Lewis, 1992). With this framework in mind, certain cognitive capacities should theoretically be in place and potentially contribute to children's ability to develop a concern for the evaluation of others (Botto & Rochat, 2019).

Although studies have yet to investigate which cognitive capacities might be necessary for EAP, there are some theoretical and empirical evidence to suggest that three specific social-cognitive capacities may be related to the emergence and development of EAP: an objectified sense of self, theory of mind (ToM), and normativity. I elaborate on each of these cognitive abilities below.

Objectified Sense of Self

A critical development observed toward the end of the second year is passing the mirror mark test. Originally studied by Amsterdam (1972), the mirror mark test probes for mirror self-recognition in young children, which has become the litmus test for explicit (i.e., conceptual) self-awareness (Bard, Todd, Bernier, Love, & Leavens, 2006; Gallup, 1970; Lewis & Brooks-Gunn, 1979; Lewis et al., 1989; Rochat, 2003). In this paradigm, the experimenter surreptitiously marks the infant with rouge and then shows them their reflection in the mirror. The idea is that if children have a self-concept (i.e., an idea of what they look like and who they are), then they will notice that the mark seen on the reflection is actually a mark on themselves and proceed to remove it.

While the mirror mark test has been widely used to test for a self-concept in the second year (18-21 months), its interpretation is controversial. A lean interpretation would suggest that the mirror mark test simply shows that the child can match their own kinesthetic image to their reflection (Mitchell, 1997; Suddendorf & Butler, 2013). Indeed, several other mammals have passed a version of the mirror mark test, including great apes and elephants (Anderson & Gallup, 2015; Plotnik, de Waal & Reiss, 2006). However, a richer interpretation of the mirror mark test posits that children who pass it, understand the self to be an entity that is not only perceivable to the self, but importantly, perceivable to others. This recognized self in the mirror would thus be a

public self that could be *potentially* evaluated by others, therefore the expression of an "objectified" sense of self (Rochat, 2013).

Support of this richer interpretation lies in children's display of self-conscious emotions, including embarrassment, when passing the mirror mark test. Displaying embarrassment is critical for understanding the psychological state of 18- to 21-month-olds, because unlike other emotions, embarrassment occurs almost exclusively in the presence of a real or imagined audience. When highlighting the difference between guilt, shame and embarrassment, Tangney et al. (1996) state that embarrassment entails "a sense of exposure and a heightened concern for others' judgments of the self" (p.1263). In other words, embarrassment places an emphasis on the audience, and how the audience might respond to ones' behavior. Interestingly, 18- to 24-month-olds who pass the mirror mark test, are also more likely to display embarrassment in situations that might elicit potential evaluations, such as when asked to dance or when they were given compliments (Lewis et al., 1989).

With the emergence of mirror-self recognition and self-conscious emotions, 18- to 21-month-old children also begin to use personal pronouns referring to the self, such as "me" or "mine" as opposed to "yours" (Lewis & Ramsay, 2004), and adjectives ("pretty" or "yucky"; Stipek, Gralinski & Kopp, 1990). That is, children begin to describe themselves to others, and use words that are relative to something else or relative to another (e.g. mine instead of yours; pretty instead of ugly). Together, this evidence suggests that when children pass the mirror mark test, they do not simply have a solipsistic representation of themselves; instead, they also have an objectified view of themselves that would include the concept of what they look like in relation to others.

With this definition in mind, a potential pre-requisite of EAP could be the development of an objectified sense of self. In theory, to develop EAP, children should understand that they are an observable (objectified) entity in the environment that can be perceived and therefore be potentially evaluated by others. To probe this hypothesis, the present study measured an objectified sense of self in toddlers using mirror-self recognition and personal pronoun use, to explore whether this cognitive ability might predict the emergence of evaluative audience perception.

Theory of Mind

Theory of mind — which is the ability to construe another's mental states, including their perspective, desires, goals, and beliefs — is a critical dimension of social cognition and social functioning. Indeed, individuals with theory of mind deficits, including individuals with autism, have trouble communicating with others, establishing social relationships, and understanding others' behavior (Baron-Cohen, 2000). Because of the key role that theory of mind plays in social development, much research has investigated how children begin to understand others' mental states (i.e., when they develop a capacity for theory of mind; Wellman & Liu, 2004). While the bulk of research has primarily focused on when children begin to understand that others can hold false beliefs (Flavell, 1999; Wimmer & Perner, 1983), research has recently begun to uncover its antecedents. These include infants' ability to perceive others as intentional agents, their understanding of pointing, and their understanding that others can have desires that differ from their own.

The pioneering work of Woodward (1998) indicated that infants conceive intentions in others by 6 months. Using a violation of expectation paradigm, Woodward tested infants' understanding of intentional actions by familiarizing them to either a human hand or a robot

hand reaching repeatedly for one of two objects (e.g., a ball instead of a teddy bear, counterbalanced across participants). This scenario established that the hand's goal was to reach for that particular object. After this familiarization phase, the objects changed position and the infants were then shown two test trials: in one of the test trials, the agent or robot hand remained consistent with its goal and reached for the ball, even though the ball's position — and thus the previous trajectory of the hand — had changed. In the other test trial, the human or robot hand followed the same trajectory, but consequently reached for the other toy. Results indicated that 6-mont-old infants reliably expected a human, but not a robot's hand, to behave according to their previously established goals. That is, they looked longer (indicating a violation of their expectation) when the hand would follow the same trajectory but reached for the teddy bear, because they expected that hand to continue following its goal to reach for the ball.

Using a more explicit method, studies have also shown that around 12 months, infants will respond appropriately to another's pointing gesture, looking toward the object and giving that person the object that they are requesting (Carpenter, Nagell, & Tomasello, 1998). This capacity to understand another's' imperative pointing further indicates that infants construe another's pointing as meaning that the person desires that object (but see D'Entremont & Seamans, 2007 for an alternative interpretation). Converging evidence support this interpretation, demonstrating that between 6 to 12 months, infants see others as intentional agents with goals and desires (see Baillargeon, Scott, & Bian, 2016 for a review).

By 15 months, children's understanding of others' desires becomes increasingly sophisticated. Repacholi and Gopnik (1997) developed the Desire Task to explore when children begin to understand that others' desires can differ from that of their own, the first step on Wellman & Liu's canonical theory of mind scale (Wellman & Liu, 2004). The authors presented

two groups of 14- and 18-month-olds with two different types of food: crackers and broccoli. In one group, the experimenter displayed a positive response to tasting the crackers, the food that most children preferred. In the other group, the experimenter expressed a positive response to tasting the broccoli, a food that the children did not prefer. After this demonstration, the experimenter would push both foods toward the child and hold her palm up to request one of the foods. The authors reasoned that if children understood that others can have different desires than that of their own, and that others' expressions towards an object match their internal desires, children should give the experimenter the food that had received a positive response, regardless of the child's own food preference. The authors found that by 18 months, toddlers would reliably give the experimenter her preferred food, suggesting that infants not only understand that others can have desires that differ from that of their own, but also that desires and opinions are reflected in behavior.

It is important to note that, on their own, the aforementioned studies do not necessarily provide evidence for a theory of mind in infancy; it is possible that these tasks are driven by associations, and thus children need not infer mental states in these tasks. However, various longitudinal studies show that infants' performance on these tasks (i.e., diverse desire, pointing) predict performance in more explicit theory of mind tasks at 3 to 4 years of age. Colonnesi and colleagues (2008) found that understanding another's imperative pointing at 12 months predicted perception understanding at 3 years of age, even after controlling for verbal IQ (Colonnesi, Rieffe, Koops, & Perucchini, 2008). Another longitudinal study echoed this finding, showing that the understanding of imperative pointing is related to the development of theory of mind (Phillips, Wellman, & Spelke, 2002). Numerous other studies have also shown that implicit measures using anticipatory looking to index goal and desire understanding in infancy also

predict theory of mind understanding at 4 years of age (Sodian et al., 2016; Thoermer et al., 2012; Wellman, LopezDuran, LaBounty, Hamilton, 2008). Such results indicate that, at the very least, these abilities emerging early in the first two years are rudiments of a theory of mind.

Lastly, research on the relation between children's theory of mind capabilities and their likelihood to perform in front of an audience suggest that theory of mind may be important for the development of EAP. Chaplin and Norton (2015) explored the relation between performance in theory of mind tasks and children's propensity to perform in front of others. The researchers hypothesized that because theory of mind involves being sensitive to others' beliefs and adopting their perspectives, children performing better in theory of mind tasks might also be less likely to perform in the presence of an audience to avoid social evaluations. To probe this hypothesis, the researchers tested 3- to 12-year-olds in three theory of mind tasks and asked the children to choose from one of two options: dancing (performing) or coloring (not performing). Results showed that the better the child performed in theory of mind tasks, the less likely they were to perform in front of the experimenter. In other words, children's developing concern for how others might evaluate them was related to their theory of mind abilities.

Likewise, research shows a link between theory of mind abilities and children's propensity to cover-up a transgression as a possible attempt to manage their public image and reputation (Talwar & Lee, 2008). For example, the tendency of 2- and 3-year-old children to engage in intentional deception (i.e., to lie), positively correlates with their performance on executive function and theory of mind tasks (Evans & Lee, 2013; Talwar & lee, 2008). One interpretation of this finding is that young children's concern with how others might evaluate them (i.e., EAP), as indexed here by their attempt to cover up a transgression, could be tied to

mental state understanding. Together, these studies provide indirect evidence for the link between theory of mind and the concern for potential social evaluations.

In sum, infants early in the second year show that they understand that others have goals, intentions and desires. Not only do they perceive others' behavior as intentional and goal directed, they also understand that others might hold different opinions and desires from their own. In theory, this ability would be a prerequisite for the development of evaluative audience perception. That is, for children to exhibit EAP, they should see others as evaluators, and understand that others can experience and evaluate things in the world in ways that might differ from their own experience[2]. To explore the relation between theory of mind and EAP, we tested toddlers in tasks aimed to measure an implicit theory of mind, including an imperative pointing task and Repacholi & Gopnik's (1997) diverse desires task.

Normativity

Researchers have argued that children younger than 3 years of age lack the cognitive capacities necessary to perceive others as evaluators of their own behavior. For example, Lewis (1992) posits that children must first develop sensitivity to norms, which are thought to develop in the preschool years, to see others as evaluators of the self. Norms are defined as the collective values within a society ("what ought to be"), or a standard ("what is;" Kallgren, Reno & Cialdini 2000, p. 1002). Echoing the presumed relation between norm understanding and sensitivity to others' evaluations, Rakoczy and Schmidt (2012) state:

> Norms set standards of correctness and appropriateness, thus guiding participants' actions in social practice and serving as reasons for justifications and as grounds for

[2] I'd like to further highlight the important difference between theory of mind and evaluative audience perception. Theory of mind simply entails the appreciation of mental states, including desires, beliefs, and intentions. Evaluative audience perception goes a step beyond a theory of mind, because it entails an appreciation for the valence of these mental states (positive or negative) and how these mental states relate to the *self*.

> critique… This means that a given norm can figure as both a reason for acting and grounds for evaluating and criticizing others' acts. (p. 17)

Thus, becoming normative means that the child can recognize others' standards and values (what ought to be), and that these values can serve as a basis for potential evaluations. Although some theorists have proposed that normativity is not evident until the age of three (Lewis, 1992), Kagan's research on norm understanding and Stipek's studies on self-evaluation suggests that normativity is evident by 24 months (Kagan, 1981; Stipek, Gralinski, & Kopp, 1990; Stipek et al., 1992). In his research, Kagan (1981) found that by 18 months, children can readily identify standards. For example, in his study, 18-month-olds would point out discrepancies with objects when they were misplaced or broken, showing they had a standard of how certain objects ought to be. From 15 to 18 months, infants would also display distress when an experimenter demonstrated a skill, such as feeding a stuffed animal, that they considered to be above their own competence. In particular, when the experimenter showed a skill that was above the child's perceived competence, the child would avoid imitating the skill and cry or cling to their mother (Kagan, 1984). Kagan interpreted the child's reaction as an indication that children could evaluate their own competence in relation to standards, and distress in these situations signaled "anxiety to possible task failure" (Kagan, 1984, p. 60).

Further empirical support for Kagan's interpretation comes from Stipek et al.'s (1992) seminal work on self-evaluation. Self-evaluation, which is evident by 24 months, requires a judgement of one's behavior or competency based on norms (Stipek et al., 1992).[3] In one specific study, the researchers showed 24 to 42-month-old children a clear fail/success task

[3] Note that self-evaluation differs from being sensitive to others' evaluations — self-evaluation does not need the perceived evaluative presence of others (i.e., an evaluative audience perception). As an illustration, a child can look at a picture they drew and evaluate their own skills at drawing (self-evaluation) without considering how others will perceive his or her drawing skills (evaluative audience perception).

where the experimenter asked the child to build a tower within a certain time frame. The researchers were interested in children's manifestations of self-conscious emotions in situations where the child failed because, if children could self-evaluate their own outcome in a task (i.e., whether they failed or succeeded in building the tower), then the child should display different emotions (i.e., happy vs. embarrassed) depending on the outcome. Indeed, the data showed that by 24 months, children evaluated their outcome differentially, displaying embarrassment and gaze aversion when their outcome was a "failure," but smiling and looking at the experimenter when succeeding at constructing the tower. Importantly, this emotional distinction was absent when the experimenter, as opposed to the child, failed or succeeded at building the tower, suggesting that the results did not simply indicate reactions to building the tower. Interestingly, the data contradicted the researchers' initial hypothesis as, like Lewis, they did not expect children younger than 3 years to show self-evaluative emotions. The authors state:

> Although we had not expected to see avoidance behaviors in children younger than 42 months, no age differences were in fact observed in any of these behaviors. Rather, attempts to avoid eye or face-to-face contact with the experimenter by turning either the head or the whole body away were observed more frequently in the failure condition among children in all age groups, and body posture was, on average, more "closed" following failure than following success. (p. 58)

Going against their original hypothesis, even the youngest of children showed that they were able to evaluate their own competence at making the tower. This suggests that children understood the norm of success versus failure, indicating that sensitivity to norms is indeed evident by a child's second birthday (Stipek et al., 1992).

In addition to being able to self-evaluate, children also begin to use evaluative words, such as pretty, little, or good, by 24 months. Regarding the meaning and implications of self-evaluative words, Stipek, Gralinski and Kopp (1990) state

> Self-description and self-evaluation precede, to some degree, emotional responses to wrongdoing, in part because children must understand that adult disapproval is directed at them, not at the act or the consequences of the act. Describing the self with evaluative language may be evidence that the child understands that he or she has not only a physically distinct self but also a self that can be evaluated positively or negatively. (p. 976).

Lastly, the tendency toward overimitation — the overly faithful reproduction of an action — further indicates normative understanding in the second year, because it shows that infants are attuned to the way things ought to be done. For instance, by 14 months, children will meticulously reproduce an action, even if such action is blatantly inefficient or causally irrelevant to a particular goal (Call, Carpenter, & Tomasello, 2005; Clay & Tennie, 2018). Researchers consider overimitation an important developmental milestone not only because it entails normativity and the need to understand the way things ought to be done, but also because it serves as a pro-active mark of affiliation, or in other words, “I am part of the group” (Over & Carpenter, 2012; Nielsen & Blank, 2011; see Hoehl et al, 2019 for review). This is evident in the fact that toddlers, for example, will overimitate an action significantly more if the adult who did the overimitation is present, highlighting the role of the social partner in overimitation (Nielsen & Blank, 2011). Fourteen- to 18-month-olds will also overimitate an actor if the action was intentional (Carpenter, Akhtar & Tomasello, 1998; Meltzoff, 1995). Thus, older infants are

increasingly selective in their overimitation tendency; they not only pay close attention and reproduce another's' action, but also consider the context and the actor.

In all, evidence points to the fact that children show signs of normative understanding by the second year. Like Lewis states, normativity should indeed contribute to children's ability to see others as evaluators, since children need to understand that there are shared expectations and values placed on certain behaviors that could serve as the basis for potential evaluations. To test this hypothesis, we had children participate in normativity tasks, including overimitating arbitrary actions on a toy and following a particular color sequence when building a tower, to investigate the relation between normative understanding and evaluative audience perception.

Summary

Based on the literature reviewed above, as well as the two defining components of evaluative audience perception, children should have at least three specific social-cognitive abilities for EAP to emerge: First, children need to construe themselves as an entity that others can see and therefore potentially evaluate (objectified sense of self). This ability is evident toward the end of the second year, as indexed by the passing of the mirror mark test, use of personal pronouns, and self-conscious emotions. Second, children should also have a basic understanding that others can perceive and evaluate things in the environment if they are to assume that others can evaluate their behavior (theory of mind). Numerous studies show that between the ages of 12-15 months, children develop rudiments of a theory of mind, construing others as intentional agents with desires that differ from that of the child. Lastly, children should also have normative understanding, as norms serve as the basis for potential evaluations (normativity). Children's sensitivity to norms and standards, as well as their tendency to overimitate, is also evident by 24 months. On the whole, the combination of these three social-

cognitive abilities, which emerge by the child's second birthday, would be pre-requisites for the emergence of evaluative audience perception.

The Present Study

Recent studies indicate that children begin to develop evaluative audience perception by 24 months (Botto & Rochat, 2018; Over & Carpenter, 2009; Rochat, Broesch & Jayne, 2012). However, we know surprisingly little about how this ability emerges and develops in early childhood (Silver & Shaw, 2018). As discussed in the previous section, there is some theoretical and empirical evidence to suggest that particular social-cognitive abilities, including an objectified sense of self, theory of mind, and normativity, might be potential building-blocks for the emergence and development of EAP. However, this conjecture has not been empirically tested. As a first step to address this gap in the literature, the primary goal of the present study was to explore whether the combination of an objectified sense of self, theory of mind, and normativity are related to the emergence of evaluative audience perception in 14- to 27-month-old-toddlers. Specifically, we wanted to test the idea that, independent of other factors, such as the child's temperament, language fluency, and general cognitive competency, children's manifestation of these three social-cognitive competencies would be the best predictor of EAP.

As a measure of evaluative audience perception, we used Botto & Rochat's (2018) robot task, which measures children's propensity to strategically modify their behavior depending on the experimenter's attention and her previously demonstrated values. In particular, we measured children's choice to activate a toy-robot via a remote control in two different situations: in one situation, two experimenters expressed opposing opinions ('yay' vs. 'yuck') toward pressing the remote and activating the toy robot. In the other situation, one experimenter expressed a positive and negative value towards pressing two different remotes. The purpose of incorporating the two

robot tasks was to not only have a more accurate measure of the emergence of EAP, but also to test the reliability of these two EAP tasks within participants.

To measure an objectified sense of self, theory of mind, and normativity, we chose six specific tasks (two for each social-cognitive ability) that met three important criteria: The first is that each task was age appropriate and yielded enough variability (Amsterdam, 1972; Call, Carpenter & Tomasello, 1998; Lewis & Ramsay, 2004; Nielsen, 2006; Repacholi & Gopnik, 1997). The second is that, within each social-cognitive ability, the two tasks measured the same construct (Colonnesi et al., 2008; Hilbrink et al., 2015; Kenward, 2012; Lewis & Ramsay, 2004; Thoermer et al., 2012). For example, the mirror mark test and personal pronoun use have been used to index an objectified sense of self and the two tasks have been shown to be related (Lewis & Ramsay, 2004; Stipek et al., 1992). The last criterion was that the chosen tasks predict performance within that particular social-cognitive ability over the course of development. For example, several studies indicate that imperative pointing understanding at 12 months is predictive of theory of mind understanding at 4 years, (Carpenter, Akhtar & Tomasello, 198l; Colonnesi et al., 2008) This provides further evidence for the tasks' validity (Kenward et al., 2011; Kristen-Antonow et al., 2015; Sodian el al., 2016).

Based on these three criteria, we measured an objectified sense of self by testing for mirror self-recognition and asking parents to document the child's use of personal pronouns. As a proxy to children's theory of mind ability, we measured children's imperative pointing understanding and their understanding of another's desires using Repacholi & Gopnik's (1997) desires task. Lastly, we measured children's normativity by measuring their norm sensitivity and their overimitation tendencies. To measure norm sensitivity, we adapted Clegg & Legare's (2016) necklace task, which measures children's propensity to reproduce a particular color

sequence previously demonstrated by the experimenter. Due to our sample's age-range, we used an age-appropriate toy tower instead of a necklace (see Method). Following the literature, we measured overimitation by documenting whether children would faithfully reproduce a modelled arbitrary action (i.e., banging toy kaleidoscope three times before looking through it).

In addition to measuring the three social-cognitive competencies, we also measured children's language and spatial ability, their temperament, and the caregiver's level of self-consciousness. The reason for including these additional measures were twofold: first, we wanted to explore whether these other factors might also be related to the emergence of EAP. For instance, it is reasonable to think that children's level of shyness or inhibitory control might be related to their button-pressing behavior in the robot task. Indeed, if children's inhibitory control is low, then they might not show EAP because their ability to be strategic in their button-pressing behavior would be diminished. Along the same vein, research has shown that parental characteristics and behavior can affect the development of social-cognitive phenomena (Moeller & Schick, 2006; Newton et al., 2014; Vaughan et al., 2003). Therefore, the caregiver's own level of self-consciousness could be related to children's expression of EAP.

The second reason for including these measures was to further probe the idea that the combination of the three-specific social-cognitive abilities proposed here is the best predictor of EAP. If this is the case, then we would not expect other more general cognitive abilities, such as language or spatial ability, to be significant predictors of EAP. Therefore, these measures not only allowed us to have a more comprehensive picture of factors that might be related to EAP's emergence, but it also served as control measures to better test our hypotheses.

To account for these factors, parents completed questionnaires that measured their public self-consciousness, the child's temperament, and the child's vocabulary competency. To have a

more comprehensive measure of temperament, we also coded for behavioral inhibition, which is a child's propensity to be withdrawn or slow to warm-up in new situations (Kagan et al., 1984). Lastly, we adapted Huttenlocher et al.'s (1994) spatial task, which measures children's ability to find a hidden toy in a three-dimensional space, as a measure of a more general cognitive ability. On the whole, the aim of this study was to serve as a critical first step to explore potential social-cognitive pre-requisites to EAP.

Hypotheses

Based on the literature regarding possible links between an objectified sense of self, theory of mind, normativity and EAP, as well as our previous work on EAP, we had three working hypotheses. The first hypothesis, which is the crux of this dissertation, is that the combination of an objectified sense of self, theory of mind and normativity would predict the emergence of evaluative audience perception in 14- to 27-month-olds. In particular, we expected children who displayed all three social-cognitive abilities to also be more likely to strategically modify their button-pressing behavior across conditions of the robot task. Our second hypothesis was that, as in our previous studies (Botto & Rochat, 2018), age would not be a significant predictor of EAP as indexed by button-pressing behavior in the robot task. Lastly, and in line with our first hypothesis, we did not expect other more general cognitive abilities, such as children's spatial ability, to predict EAP. That is, we did not expect children's performance in the spatial task to predict strategic button pressing in the robot task. Together, these three hypotheses aimed to test the idea that, even when accounting for maturation (age) or general cognitive competency (spatial ability), the combination of the three proposed social-cognitive abilities would be the best predictor of EAP in toddlers.

Method

Participants

Fifty-nine 14- to 27-month-old (*M* = 20.8, *SD* = 4.4; 37 females) healthy children living in a large urban city participated in this study. All participants were recruited from the Emory University Child Study Center database. An additional 4 children participated but were excluded from analysis due to fussiness (n = 3) or parental interference (n = 1). Sample size was pre-determined using G*Power 3.1 analysis, which determined that 60 participants were needed to achieve 85% power and a medium effect size (Faul, Erdfelder, Lang, & Buchner, 2007). The sample consisted of 61.7% Caucasian, 15% mixed-race, 8.3% African American, 6.7% Asian, and 1.7% Latino children. Approximately 85% of parents reported having a college degree. All participants attended one testing session with two experimenters (E1 and E2) and participated in all of the tasks. This study was approved by the Emory University Internal Review Board under the project name Origins of the Intentional Stance (Study Number IRB00041083).

Measures and Procedure

General Procedure and Experimental Room set-up

All testing sessions were video recorded for coding purposes using a total of six small video cameras from different angles. The testing room was divided into two sections by a curtain with three cameras on each side (See Figure 1). For the first part of the session, one camera faced E1, the child and caregiver while they were sitting on the floor, a second camera was positioned for an aerial view, and the third camera was placed behind a two-way mirror. For the second part of the session, one camera was placed on a table behind the child, a second camera was positioned behind the experimenters to the left of them, and the third camera was positioned behind the experimenters between them. These angles gave an unobstructed view of the child's

behavior, as well as a close-up view on the child's face to code gaze and behavior. Both of the cameras that were located behind the experimenters did not have the experimenters in the frame to allow the coders to be blind to conditions. To control for the appearance of E1 and E2, both researchers were Caucasian brunettes, wore their hair up, and wore either a blue or orange scrub (color of the top was counterbalanced across participants).

Before the testing session began, the experimenters interacted with the participant outside of the testing room for five minutes to get them comfortable and familiarized with the experimenters. This five-minute warm-up period also served as an opportunity to code for children's level of behavioral inhibition (see Coding below). After the warm-up period, the experimenters brought the caregiver and the child into the testing room. To prevent parental interference, the caregiver was given a pair of sunglasses to wear at the beginning of the study and was asked to remain quiet and neutral for the duration of the testing session.

The caregiver was instructed to sit on a marked spot on the floor with the child in their lap facing E1. Once the child sat down, E1 stated "You are so cute!" and surreptitiously marked the child with odorless purple eyeshadow to use for the mirror mark test later in the session. Marking the child at the beginning of the session minimized the chance that the child would notice being marked and allowed coders to document any self-directed behavior that would indicate the child noticed being marked (see Coding for details).

Following the short warm-up period, all participants completed the seven cognitive tasks in the same sequential order[4] (See Figure 2). The measures and procedures for each task is described below, following the order in which they were presented to the child participants

[4] The order of tasks was decided based on piloting to optimize the child's engagement. Further, there was no evidence in the literature to suggest that the order of tasks would influence performance across tasks.

during the testing session. Details for passing criteria for each task can be found in the Coding section (see Table 1 for summary of measures and coding criteria).

Imperative Pointing Task (theory of mind)

The first task presented to children was the imperative pointing task, a theory of mind task that measures children's understanding of intentionality based on the experimenter's point and gaze (Carpenter, Nagell, & Tomasello, 1998; Colonessi et al., 2008). To begin the task, E1 presented the child with a stuffed lion and a toy car to engage the child. After the child engaged with the toys for two minutes, E1 turned her head to the left, pointed to a 7.62 cm pink bendable-vinyl toy pig that was 180 cm away from her, and exclaimed, "Look! What's that?" After the child looked at the toy pig and back at the experimenter, the experimenter opened her palm and tried to reach toward the toy. Passing the imperative pointing task meant that the child inferred E1's desire for the pig based on her gazing and pointing, and thus the child should respond to E1's gesture by retrieving the toy pig. If children did not respond initially, the experimenter repeated her hand gestured a maximum of two more times, and asked, "Can I have it?" In cases where the child gave the pig to the experimenter, E1 thanked the child and put the toy away.

Diverse Desires Task (theory of mind)

Following the imperative pointing task, children participated in Repacholi & Gopnik's (1997) diverse desires task, the second theory of mind task. In this task, children's understanding of another's desire is measured by probing whether children will give the experimenter her preferred food, even if that food is not preferred by the child. Unlike Repacholi & Gopnik's original paradigm, we opted to use toy foods as opposed to real food, because research suggests that perceptual features (i.e., the fact that real food might be tempting to eat) may interfere with task performance. Therefore, using symbolic stimuli would boost task performance in the desires

task (Carlson, Davos & Leach, 2005). For the two food options, a 7.6 cm toy cookie and a 7 cm toy broccoli were placed in a 26 x 15.25 cm transparent box with no lid. The box allowed for the two foods options to be the same distance away from each other as well as the child during the task (placement of broccoli and cookie was counterbalanced across participants). To show children the food options, E1 picked-up one of the toy foods (order was counterbalanced across participants) and said "Look! There is broccoli/cookie" while pretending to eat the food and then repeated the process with the other toy food. After demonstrating both toy foods, E1 placed to box near the child and asked, "Which one do you like? Pick one!" After the child picked his or her preferred food, E1 stated, "Let me show you which one I like. I like broccoli/cookie!" E1 picked the toy food that was not selected by the child and pretended to eat it while saying "mmm I love broccoli/cookie!" After, the experimenter placed her preferred food back in the box, she grabbed the child's preferred food, and stated, "I *don't* like broccoli/cookie! Yuck, broccoli/cookie!" After this demonstration, E1 pushed the box toward the child and stated "I'm hungry. Will you give me one to eat?" while rubbing her stomach with the left hand and extending her right palm toward the child. If the child did not respond, E1 repeated, "I'm hungry. Will you give me one to eat?" a maximum of two times. Passing the desires task entailed the child giving E1 her preferred food. After the child made their selection, E1 put the foods away.

Mirror Mark Task (objectified sense of self)

To test for mirror self-recognition, children were surreptitiously marked on their nose with a bright purple, odorless play eyeshadow at the beginning of the study, and were later shown their reflection on a 91.5 x 51 cm two-way mirror. The two-way mirror allowed us to place a hidden camera directly behind it, allowing coders to have an unobstructed view of the child's reaction (i.e., touching mark on face) when looking directly at their reflection. So that

children could not see their reflection prior to the mirror mark test, the mirror was covered with a black cloth. When it was time for the mirror-mark test, E1 showed the child a yellow ball and placed it in front of the covered mirror to entice the child to stand in front of the mirror. E1 then asked the child, “Can you come get the ball?” Once the child reached for the ball in front of the covered mirror, E2 lifted the black cloth to reveal the two-way mirror. After 30 seconds, E1 pointed to the child’s reflection and as a prompt said, “Who is that?” a maximum of three times. The task lasted a maximum of three minutes, or until the child touched the mark on their face, which signified successful mirror self-recognition.

Robot Task 1 (evaluative audience perception)

To measure evaluative audience perception, we used two versions (robot task 1 and 2) of Botto & Rochat’s (2018) robot task. The robot task measures children’s propensity to strategically modify their behavior depending on the experimenter’s attention and her previously demonstrated values. For Robot Task 1, we used a 13 x 13 cm red circular toy robot that could be activated by a remote control. The toy robot was placed inside a 30.5 x 28 cm blue box that was placed on its side, and the remote control was placed in front of the box. E1 and E2 sat side by side, opposite the child who was sitting across a table on their caregiver’s lap (see Figure 1). During the demonstration phase, E1 picked-up the remote, said “Look!” and pressed the remote. After the robot moved, E1 offered a positive feedback by smiling and saying “Yay! The toy moved!” This was repeated twice. E2 then picked up the remote and repeated the same process as E1, but expressed negative feedback after the toy moved, frowning and stating, “Yuck! The toy moved!” The full demonstration was repeated twice.

After the full demonstration, E1 and E2 placed their hands on the remote and pushed it towards the child simultaneously while saying, “Your turn!” to invite the child to interact with

the remote. E1 (positive feedback) proceeded to turn around 180 degrees while E2 (negative feedback) faced the child for 30 seconds (negative-attentive condition). After 30 seconds, E1 and E2 switched so that E1 (positive feedback) was facing the child and E2 (negative feedback) turned her back (positive-attentive condition). In this way, the child always had an experimenter watching while they interacted with the remote, but the type of experimenter (positive versus negative feedback) changed between conditions. The feedback that E1 and E2 gave, the order the negative and positive feedback were presented to the child, and which experimenter turned around first were all counterbalanced across participants. The amount of button-presses across both conditions (negative-attentive and positive-attentive) was later coded by research assistants using an event recording software (see Coding below).

Overimitation Task (normativity)

After children completed the first robot task, we measured the child's propensity to overimitate arbitrary actions on an object. For the object, we chose a 15.2 cm blue toy kaleidoscope (i.e., cylinder with beads on one end). To model the arbitrary action, E1 picked-up the kaleidoscope, banged it slowly on the table 3 times (arbitrary action) and then proceeded to look through it while saying "wow!" This sequence was repeated three time. After the demonstration phase, E1 placed the kaleidoscope in front of the child and said, "Your turn!" The child was allowed to interact with the kaleidoscope for 60 seconds. Children's propensity to bang the kaleidoscope at least once before looking through it counted as overimitation.

Tower Task (normativity)

As a second measure of normativity, we used a novel paradigm called the 'tower task'. Adapted from Clegg and Legare's (2016) necklace-making task, the tower task aimed to measure faithful reproduction of stacking colorful rings based on a size and color sequence previously

shown to the child. Because we used a stacking-ring toy that was commercially available, we replaced the typical conical pole for a straight 28 cm pole to ensure that the child would not stack the rings based on fit, but rather based on the order shown by the experimenter. At the start of the demonstration phase, E1 drew the child's attention to the correctly stacked tower. E1 then removed the rings form the pole and placed the rings in a randomized order in a clear box and stated, "Let's see if we can put this back together." E1 proceeded to pick-up the largest ring, show it to the child, and said, "We get the blue one," placing it on tower. E1 then repeated the same motion with the second largest ring. To highlight that there was a particular order, E1 picked- up the smallest ring, showed it to the child and stated, "do we get the red one?" After placing it on the pole, she exclaimed "Nope! That's not right," while shaking her head side to side. E1 then proceeded to slowly build the tower from smallest to largest, carefully showing the child each ring before placing it on the tower.

After E1 completed the demonstration of how to properly stack the ring tower, E1 pointed to the finished tower, said, "Look!" in order to again draw the child's attention to the finished tower and then said while clapping, "Yay! I did it right!" While pointing to the child, E1 said, "Can you do it?" E1 then placed the rings back into the clear box in a randomized order and then pushed the box and the empty tower pole closer to the child stating, "Your turn!" If the child failed to follow the correct order sequence, the experimenter shook her head from side to side and stated "that's not right" after the child had finished stacking all of the rings. The experimenter then proceeded to do the demonstration again, giving the child another opportunity to follow the correct size and color sequence. Children who followed the correct color sequence on either trial (first or second time) counted as passing the tower task.

Robot Task 2 (evaluative audience perception)

Following the tower task, children participated in the second robot task (one experimenter and two remotes). The same materials that were used for Robot Task 2, with the exception that this task used *two* 13.5 x 9 x 2 cm remote controls with one button on either end. To clearly distinguish the two remotes and to expose only one button, each remote was placed in either a blue or orange sock (colors were counterbalanced). When activated, each remote caused different robot motions (moving forward versus spinning in circles). We incorporated these different robot motions to entice the child to explore both remotes. Both remotes were placed in a 26 x 15.25 x 7.6 cm transparent box to the right of the robot. For the demonstration phase, E1 picked up one remote, pushed its button, and gave positive feedback ("Wow! Isn't that great?") and smiled, or negative feedback ("Uh oh! Oops, oh no!") and frowned, after setting the robot in motion. After pushing the first remote twice, the experimenter then picked up and demonstrated the effect of pressing the button of the other remote, with accompanying positive or negative feedback. The experimenter gave this feedback only during the demonstration phase.

After the demonstration phase, the experimenter smiled and pushed the tray with the two remotes toward the child and said, "Your turn!" The experimenter then either turned 180° and pretended to read a magazine for 30 seconds (inattentive condition) or looked at the child with a neutral face for 30 seconds (attentive condition), order counterbalanced. The color of remote, movement of the robot, type of feedback, as well as left–right location of the remote on the tray was counterbalanced across participants. Button-pressing behavior between the two remotes across attentive and inattentive conditions were coded by research assistants using an event-recording software (see Coding section below).

Spatial Task (general cognitive ability)

To measure the specificity of the potential relation between the three social-cognitive abilities and EAP, children participated in a novel task that measured spatial ability at the end of the testing session. Adapted from Huttenlocher, Newcombe, and Sandberg (1994), the spatial task measures spatial awareness in 16- to 24-month-olds. In their original task, the authors used a 152 x 183 cm sandbox and hid a small toy in predetermined locations. A child's spatial ability was indexed by children's search accuracy (i.e., proximity of child's search location to the actual location of the hidden toy). Because of space constraints, we used a string of 32 blue and white pillboxes, which were connected to make an 83.8 x 5 x 2.5 cm chain. This chain was placed on a 105 x 60 x 53 cm table so that the child could freely search for a 5-cm rubber toy turtle.

To ensure task comprehension, the task began with 3 practice trials. In the practice trials, E1 would show the child the toy turtle, introduced as "Mr. Turtle" to the child, and place the turtle inside one of the pill boxes. To ensure that the child was watching where E1 hid Mr. Turtle, E1 would tap the lid of the pill box three times once the toy was inside. Then, E1 would ask the child to find Mr. Turtle. This sequence was repeated a total of three times, with the turtle being hidden in three separate pre-determined spots (in the center pill box and at each extremity of the chain). Children who successfully found the turtle in all three spots proceeded to the 5 test trials. For the 5 test trials, the experimenter hid the toy turtle in 5 pre-determined, randomized locations before asking the child to locate the hidden toy. The difference between the practice trials and the test trials, was that after E1 showed the child where the turtle was hidden, she proceeded to distract the child by asking them to look at their caregiver. After the child looked at their caregiver, E1 asked them to find Mr. Turtle. The child's spatial ability was measured by their search accuracy across all five trials, giving each child a single continuous spatial score.

Parental Questionnaires

After the testing session, the caregiver filled out four questionnaires that measured the child's temperament, vocabulary (including use of personal pronouns), and the parent's own self-consciousness. The description of each questionnaire is below:

Demographics Questionnaire. The demographics questionnaire recorded participants' race, age, gender, number of siblings, parent's education, and whether the child attended daycare.

Early Childhood Behavior Questionnaire (ECBQ). To account for temperament, parents filled-out the Putnam, Gartstein, and Rothbart (2006) short temperament assessment form of the Early Childhood Questionnaire, a version appropriate for children age 14 to 32 months. This questionnaire asks parents to rate how much an item applied to their child in the past 2 weeks using a 7-point Likert scale (1 = *never*, 7 = *almost always*) for five different dimensions of temperament. A mean score is then calculated for each dimension. For the analysis, we were particularly interested in the dimensions of Inhibitory Control and Shyness to see whether these temperament dimensions related to children's behavior in the robot task. The shyness dimension included five items, including, "When approaching unfamiliar children playing, how often did your child watch rather than join in?" The inhibitory control dimension included six items, including "When told "no," how often did your child stop the forbidden activity?*"* Both dimensions have good internal validity (alphas = 0.78 & 0.88, respectively; Putnam et al., 2006).

McArthur-Bates Communicative Development Inventory. To measure language competency and children's personal pronoun use (I, me, mine, yours), parents completed the McArthur-Bates Communicative Development Inventory. The McArthur-Bates is a 680-item questionnaire that probes children's production and understanding of vocabulary words in nineteen semantic categories, including pronoun use, animal names and body parts. This

questionnaire is a widely used measure for speech production and comprehension in late infancy (Fenson, 2007).

The Self-Consciousness Scale-Revised Version (SCSR). The SCSR is a 22-item valid measure that probes private and public self-consciousness and social anxiety in general populations (Cronbach's Alpha = .75, .84, & .79, respectively; Scheier & Carver, 1985). The private self-consciousness score indicates an individual's level of introspection. The public-self-consciousness score measures individual's awareness of others' opinions toward the self. Lastly, the social anxiety score measures the degree to which an individual might experience anxiety in social settings. Adults are asked to rate statements using a 4-point Likert scale (0 = *not like me at all*, 3 = *a lot like me*). Sample items include, "I am concerned about what others think of me," and "I get embarrassed very easily."

Coding

All coding was completed using prerecorded videos by trained research assistants who were blind to the conditions and hypotheses. All tasks, with the exception of button-presses in the robot task and the spatial score, were coded categorically, with 1 indicating passing a task and 0 indicating not passing. Passing criteria for each task is described below:

Mirror Mark Test and Personal Pronoun Use (objectified sense of self). To ensure that children were unaware of being marked the beginning of the study, research assistants first coded for any self-directed touches within 15-seconds after the child was mark. Based on this criterion, none of the children in the study noticed being marked. To count as passing the mirror mark test, the child had to touch and/or removed the mark on their nose after seeing their reflection. Those who did not touch the mark received a score of 0. For personal pronoun use, children who used

at least two personal pronouns as indicated by the McArthur-Bates Communicative Development Inventory received a score of 1, and those who did not received a score of 0.

Imperative Pointing Task and Desires Task (theory of mind). To pass the imperative pointing task, children had to follow the gaze of the experimenter to the external object (i.e., the toy pig), look back at the experimenter, and successfully retrieve the toy pig and hand it to the experimenter. Those who did so successfully received a 1 and those who did not received a 0. For the desires task, children who gave the experimenter their correct food preference received a score of 1 and children who did not respond or give the experimenter the correct food choice received a 0.

Overimitation and Tower Task (normativity). To capture overimitation, children who banged the kaleidoscope at least once before looking through it received 1 point. Those who failed to do so received 0 points. For the tower task, children who stacked the rings in the correct color order in either trial (1 or 2) received 1 point, and those who did not received a score of 0.

Robot Task 1 and 2 (evaluative audience perception). To measure EAP, children's button presses during conditions of the robot task were recorded using Boris, an event-recorder software (Friard & Gamba, 2016). Only the button presses that successfully activated the robot counted. For Robot Task 1 (two experimenters and one remote), displaying EAP meant that the child pressed the remote significantly more if the positive experimenter was watching compared to the negative experimenter. For Robot Task 2 (one experimenter and two remotes), displaying EAP meant that the child pressed the positive remote significantly more when the experimenter was watching (attentive condition), but chose to press the negative remote significantly more when she was not watching (inattentive condition). Raw button-presses across conditions of each robot task were used as the dependent variable.

Spatial Task (general cognitive ability). The spatial task measured children's ability to successfully locate a hidden toy in a 3-dimensional space. Children's accuracy was rated by calculating the distance between the child's first searching spot and the correct spot. Because there were 5 trials, children's total spatial score consisted of their score in each trial divided by 5. In this way, lower scores indicated better performance on the spatial task.

Behavioral Inhibition (temperament). As a second measure of temperament, research assistants coded for signs of behavioral inhibition within the first five minutes of the study. Coders used Kagan et al.'s (1984) criteria for behavioral inhibition, which included apprehension to interact with the experimenter, refusal to interact with the experimenter or with the toys, freezing, such as stopping play or vocalization, and clinging to or burying themselves in the caregiver. For each of these behaviors shown within the first five minutes, the child received 1 point. For example, if the child clung to mom and also refused to interact with the experimenter when she was engaging the child, the child received a score of 2. As such, each child received a score of 0-3 for behavioral inhibition, with 3 being the highest (i.e., showing the highest level of behavioral inhibition).

Questionnaires: Language, temperament, and public self-consciousness. All questionnaires, with the exception of demographics, were coded according to the instructions specified and provided by the creators of each questionnaire. The questionnaire scores were entered either as a continuous variable (language score), or an ordinal variable (Early Childhood Behavioral Questionnaire, Self-Consciousness Scale-Revised). For the language score, coders calculated a vocabulary percentage by dividing the total number of words that the child could pronounce and understand by the total number of possible items (680). For the Early Childhood

Behavioral Questionnaire and the Self-Consciousness Scale, a score was derived for each dimension using a coding scheme provided by the authors of each questionnaire.

Scoring of Each Social-Cognitive Ability

For the child to count as manifesting a particular social-cognitive ability (i.e., an objectified sense of self, theory of mind, or normativity) the child had to pass at least one of the two tasks measuring that particular ability. For example, if the child received a score of 0 in the mirror mark test but received a score of 1 for personal pronoun use, the child was still considered as having shown an objectified sense of self and would receive a score of 1. In this way, each child received either a 0 or 1 for each social-cognitive ability depending on whether they passed at least one task within a particular social-cognitive ability. This approach prevented children from not getting "credit" if they did not pass just one of the tasks within a cognitive measure. Importantly, we were primarily interested in whether or not the child manifested a particular ability, not the extent that they showed that particular social-cognitive ability.

Reliability

After extensive training on the passing criteria for each cognitive task, a second coder coded 20% of all participants. Coder reliability was excellent across all tasks; for the six categorical measures (passing or failing in each social-cognitive task), interclass correlation coefficients ranged from .86 to .97. For the two continuous measures (button presses in robot task and search accuracy in spatial task), coders achieved good reliability (Cronbach's Alpha = .77 & .97, respectively).

Planned Analyses

Descriptive statistics for each task were obtained and a series of logistic regressions were conducted to examine whether age, gender, or language competency were related to each task.

We also used McNemar's tests, which test associations between two paired nominal data, to examine the relation between tasks within each social-cognitive ability. Normality tests and frequency distributions were also conducted to ensure normal data distribution and identify any outliers.

Preliminary analyses using mixed factorial analysis of variance (ANOVA) examined whether factors other than the child's social-cognitive competency predicted EAP. These factors included the child's age (Hypothesis 2), gender, temperament, and overall language competency, as well as the parent's level of public self-consciousness. Conducting these preliminary analyses allowed us to identify significant predictors of EAP that would need to be included as covariates in the main analysis, while also maximizing power.

For the main analyses, a mixed ANOVA was used to test the hypothesis that manifesting all three social-cognitive abilities would predict EAP (Hypothesis 1). For the purpose of this analysis, we created three social-cognitive competency groups. The three groups consisted of the number of social-cognitive abilities (i.e., objectified sense of self, theory of mind or normativity) that the child manifested (manifesting 1, 2 or all 3; see Coding for details). In this way, the independent variable for the main analysis was the child's social-cognitive competency as indexed by the three groups, and the dependent variable was the child's button-pressing behavior across conditions of the robot task. Lastly, we ran a mixed model ANOVA with performance on the spatial task as a predictor variable to test the hypothesis that the child's general cognitive competency would not be a significant predictor of EAP (Hypothesis 3). All analyses were performed on IBM SPSS 26 data analysis software.

Results

Descriptive Statistics and Relations Between Tasks:

Evaluative Audience Perception:

*Robot Task 1 and 2***.** Table 2 shows the descriptive statistics of raw button-pressing behavior for both robot tasks. For Robot Task 1 (two experimenters and one remote), 23 out of 59 children (39%) showed the effect of pressing the remote significantly more when the positive as opposed to the negative experimenter was watching. A binary logistic regression showed that neither age nor gender were related to showing the effect, $\chi^2(2) = .88$, $p = .65$. For Robot Task 2 (one experimenter and two remotes), 34 out of 59 (57.6%) children showed the effect of pressing the negative remote more when E1 was not watching compared to when she was watching. Again, neither age nor gender were related to showing the EAP effect in robot task 2, $\chi^2(2) = 2.1$, $p = .35$. To explore the relations between the two tasks, we used a McNemar's test, an analysis that examines associations between paired nominal data (Adedokun, & Burgess, 2012). Results indicated that children who showed the effect in one robot task were also significantly more likely to show the effect in the other robot task, $p = .02$ (two-sided). Because more children passed Robot Task 2 (1 experimenter and 2 remotes), and because its effect has been validated by a control study in a previous paper (Botto & Rochat, 2018), subsequent analyses used raw button presses of Robot Task 2 as the dependent variable.

Objectified Sense of Self

Mirror mark test and personal pronouns. Of 59 children, 16 (27.1%) showed mirror self-recognition and 24 (40.6%) used at least two personal pronouns (see Table 3). Consistent with the literature (Rochat, Broesch & Jayne, 2012; Zmyj, Prinz, Daum, 2013), a binary logistic regression indicated that age, but not gender predicted passing of the mirror mark test, $\chi^2(2) = 16.84$, $p < .001$, with older children being significantly more likely to pass the mirror mark test. For use of personal pronouns, parents of 24 out of 55 children (40.7%) indicated that their child

used at least two personal pronouns. The mean use of personal pronouns was 4.87 (*SD* = 4.1). A binary logistic regression showed that neither age nor gender predicted use of personal pronouns, χ^2 (2) = 5.2, *p* = .07. A binary logistic regression with passing at least one of the two objectified sense of self tasks (the mirror mark test or using personal pronouns) as the dependent variable and age, gender and total language score, revealed no significant results, χ^2 (4) = 3.4, *p* = .33. Unlike previous studies (Lewis & Ramsay, 2004), a McNemar's test indicated that children who passed the mirror mark test were not significantly more likely to use personal pronouns *p* = .134 (two-sided).

Theory of mind

Imperative pointing and diverse desires. Forty-three out of 59 (72.9%) children gave the toy to E1 after her pointing gesture. A binary logistic regression showed that age, but not gender, was related to understanding imperative pointing *b* = .4, Wald χ^2 (1)=13.98, *p* < .001, with those kids who passed the imperative pointing task being significantly older than those who did not pass. For the desires task, 31 out of 59 (52.5%) gave E1's correct food preference when requested. Neither age nor gender were related to passing the desires task, χ^2 (2) =.1.98, *p* = .38. A binary logistic regression with passing at least one of the two tasks (imperative pointing or desires task) as the predicted variable and age, gender and language scores as predictor variables revealed that only age significantly predicted theory of mind, *b* = .27, Wald χ^2 (1) = 6.1, *p* = .01. A McNemar's test showed that these two theory of mind tasks were related, with children being significantly more likely to pass the desires task if they also understood imperative pointing, *p* = .04 (two-sided)

Normativity

Overimitation and normativity. For the overimitation task, 30 (50.8%) out of 59 participants successfully overimitated, banging the toy kaleidoscope on the table before looking through it. A binary logistic regression showed that neither age nor gender were related to Overimitation, $\chi^2(2) = .21$, $p = .9$. For the normativity tower task, only 8 (13.6%) out of 59 children who completed the task successfully stacked the rings in order. A binary logistic regression showed that age, but not gender was related to successfully completing the tower task, with older kids having more success than younger kids, $b = 0.39$ Wald $\chi^2(1) = 5.2$, $p = .002$. A binary logistic regression with passing at least one of the two tasks (overimitation or tower task) as the dependent variable and age, gender and language scores revealed that these variables were not significant predictors of normativity, $\chi^2(3) = .31$, $p = .96$. To test the relation between both normativity tasks, a McNemar's test revealed that those who passed the ring test were also more likely to pass the overimitation task, $p < .001$.

Language Ability

Fifty-six out of 59 children had complete data for the McArthur-Bates Communicative Development Inventory. On average, children knew about 30% of the words listed on the questionnaire ($M = 0.3$, $SD = 0.27$; see Table 4). A linear regression indicated that age but not gender was associated with language competency, with older children understanding and uttering more vocabulary words than younger children, $F(2, 55) = 30.8$, $p < .001$, $R^2 = .54$. Data were normally distributed (skewness and kurtosis less than -2 and 2; Fields, 2009; George & Mallery, 2010).

Spatial Task (General cognitive ability)

Due to either fussiness or inability to successfully complete practice trials, 15 out of 59 (25%) of children did not completed the spatial task. One-way ANOVAs indicated that children

who did not complete the spatial task did not significantly differ from those who did complete the task in age, $F(1, 58) = 0.01$, $p = .94$ or language competency, $F(1, 55) = .86$, $p = .35$. On average, children had a mean score of 7.1 ($SD = 3.22$). A lower number indicates greater accuracy, or a smaller average distance from children's' first searching spot to correct location. The scores were normally distributed (skewness and kurtosis between -1 and 1; see Table 4). A linear regression indicated that age, gender, or language fluency did not predict performance on the spatial task, $F(3,41) = 0.27$, $p = .85$, $R^2 = .02$.

Temperament and Self-Consciousness Scale Score (Parental questionnaires)

The mean, standard deviation, and range for dimensions of interest of the temperament[5] and self-consciousness questionnaire can be found in Table 4. All data was normally distributed (skewness and kurtosis between -2 and 2; Fields, 2009; George & Mallery, 2010).

Social-Cognitive Competency Groups

As previously mentioned, three groups were created to test the hypothesis that manifesting all three social-cognitive abilities predicted EAP. The three groups consisted of the number of social-cognitive abilities the child manifested (children showing 1 or none, children showing two, and children showing all 3; see Coding and Planned Analysis for rationale and criteria). One-way analyses of variance (ANOVAs) showed that the three social-cognitive competency groups did not significantly differ in age, language scores, and total button-pressing across conditions of the robot task ($p = .8$, $p = .27$, $p = .7$, respectively). A Krustal-Wallis test

[5] To explore the relation between behavioral inhibition and the shyness and inhibitory control dimensions of the Early Childhood Behavioral Questionnaire (ECBQ), we performed Spearman Rho correlations. Results showed that our measure of behavioral inhibition and these two dimensions these were not related (r = .78, .98). This could have been because our behavioral inhibition measure was not sensitive enough to detect interindividual differences in behavioral inhibition, or because scoring for behavioral inhibition captured the behavior of the child for five minutes, whereas the ECBQ asks for the presence of certain behavior for a longer timespan. Therefore, subsequent analyses used the scores of the ECBQ as a measure of temperament, since the ECBQ has been shown to be a valid and reliable measure of shyness and inhibitory control.

revealed significant differences in the amount of shyness between children in Group 1 and Group 3, $\chi^2(2) = 6.79$, $p = .03$, with children in Group 1 having higher scores on the shyness dimension of the Early Childhood Behavioral Questionnaire. To examine whether shyness would affect button-pressing behavior when the experimenter was looking at the child (i.e., attentive condition), we compared Group 1 and Group 3's overall button-pressing in the attentive condition of the robot task. We reasoned that if differences in shyness could potentially affect how children acted in the robot task, then we would find differences in the level of remote engagement, particularly when the experimenter was watching the child. A one-way ANOVA revealed that children in Groups 1 and 3 did not differ in their button-pressing behavior in the attentive condition F (1, 56) = .63, $p = .53$, suggesting that children's level of shyness did not interfere with their overall button-pressing behavior in the robot task.

Preliminary Analyses

We first probed whether factors other than the three social-cognitive abilities we measured would be significant predictors of EAP. This preliminary analysis allowed us to identify significant predictors of EAP that would need to be factored into the model when testing our main hypothesis, while also maximizing power. A mixed factorial ANOVA with button pressing behavior between conditions in the robot task as the dependent variable, gender as a between-subjects factor, and inhibitory control, shyness, overall language fluency, and the mother's score on the public self-consciousness scale as predictor variables revealed a significant main effect of audience. Post hoc analyses indicated that children tended to press the remote significantly more when the experimenter was not looking, compared to when she was looking, F (1, 51) = 5.77, $p = .02$, $\eta_p^2 = 0.1$. However, there were no other significant findings (all $p >$

.05). Because these factors were not related to EAP, they were not included in subsequent analyses.

Age and Evaluative Audience Perception

To examine whether age was related to button-pressing behavior in the robot task, a repeated measures analysis of covariance (ANCOVA) with button-pressing behavior across conditions of the robot task as the dependent variable, and age as a predictor variable revealed no significant effects, F (1, 57) = .317, p = .57, η_p^2 = .01. Similar to findings in our previous studies (Botto & Rochat, 2018), age does not seem to be a significant predictor of evaluative audience perception.

Relation Between Social-Cognitive Competence and Evaluative Audience Perception

To directly test the hypothesis that manifesting all three social-cognitive abilities would predict EAP, we conducted a 3 (Social-Cognitive Group: 1, 2 ,3) x 2 (Remote: positive, negative) x 2(Condition: attentive, inattentive) mixed factorial analysis of variance (ANOVA), with button-pressing behavior as the dependent measure and social-cognitive group as the between-subjects factor. Results revealed a significant Cognitive Group (Group: 1, 2, 3) x Remote (positive, negative) x Condition (attentive, inattentive) interaction, F (2, 56) = 5.9, p = .005, η_p^2 = .18. Follow-up pairwise comparisons after Bonferroni corrections revealed that children in Group 3[6] were the only group to display EAP as indexed by strategic button-pressing in the robot task. Specifically, children in Group 3 activated the positive remote significantly

[6] To ensure that the results were not due to differences in sample size between the three cognitive groups, we randomly selected a subsample of participants from Group 1 and Group 2 to match the sample size of Group 3 and re-analyzed the data. Results remained consistent, yielding a marginally significant Group (1,2,3) x Remote (negative, positive) x Condition (attentive, inattentive) interaction, F (2, 27) = 3.36, p = .05, η_p^2 = 0.2. Similar to the analysis with the entire sample, post hoc analyses revealed that Group 3 was the only group button-pressing behavior, pressing the positive remote more (M = 4.6, SE =1.4) when the experimenter was watching compared to when she was not watching (M =1.6, SE =.66; p = .008), but pressing the negative remote more when the experimenter was not watching (M = 5, SE = 1.1) compared to when she was watching, (M = 1.9, SE = .72; p = .03). These findings indicate that our results cannot be explained by differences in sample sizes.

more when the experimenter was watching (M = 4.6, SE = 1.1) compared to when she was not watching (M = 1.9, SE = 0.8; p = .01), but in contrast, activated the negative remote more when she was not watching (M = 5, SE = .9) compared to when she was watching (M = 1.6, SE = 0.84; p = .04; see Figure 3). This suggests that, as hypothesized, children who displayed all three social-cognitive abilities were the ones to manifest EAP.

Relation Between Spatial Task and Evaluative Audience Perception

To test the specificity of these particular social-cognitive abilities and its relation to EAP, we also entered performance on the spatial task into the model. Specifically, we wanted to test the hypothesis that overall cognitive competency could not explain the significant relation between the three social-cognitive abilities and EAP (i.e., kids who simply perform better in any cognitive task would display EAP). As hypothesized, results indicated that spatial score did not predict EAP, F (1, 41) = 1.03, p = .32, η_p^2 = .03. This suggests that our results cannot be explained by overall cognitive competency.

Post-Hoc Analyses

To further examine whether particular social-cognitive capacities might individually predict button-pressing behavior, a series of separate mixed ANOVAs were conducted, with an objectified sense of self, normativity, and theory of mind as between-subjects factors and button-pressing behavior in Robot Task 2 as the dependent variable. Results showed that, on their own, neither an objectified sense of self, F (1, 57) = .58, p = .49, η_p^2 = .01, nor theory of mind F (1, 57) = 3.41, p = .07, η_p^2 = 0.06, predict button-pressing behavior. However, there was a significant Condition (attentive, inattentive) x Remote (positive, negative) x Normativity (passing, not passing) interaction, F (1, 57) = 4.38, p = .04, η_p^2 = 0.07. Post-hoc analyses revealed that children who displayed normativity were significantly more likely to press the

positive remote in the attentive condition (M = 3.4, SE =.52) compared to the inattentive condition (M = 2.1, SE =.46), p = .01. Note, however, that there was no significant interaction between the negative remote and audience versus no audience conditions. This indicates that, while children did choose to press the positive remote when the experimenter was watching, they did not choose the press the negative remote significantly more when the experimenter was not watching.

We also analyzed children within Group 2 (i.e., those who only showed two social-cognitive abilities) to see if any particular combination of social-cognitive abilities might predict button-pressing behavior (i.e., theory of mind and normativity vs. theory of mind and objectified sense of self, etc.) To do this, we again created three groups depending on the combination of cognitive abilities that children in Group 2 demonstrated. A mixed factorial ANOVA revealed that different combinations of cognitive abilities did not predict button pressing behavior, F (2, 24) = 3.1, p = .07, η_p^2 = 0.2. On the whole, these results provide support for the hypothesis that it is the combination of all three cognitive abilities that are the best predictor of the emergence of EAP[7].

Discussion

For the past few decades, research has documented adult's and children's concern for how others might evaluate their behavior, what we have referred to here as evaluative audience

[7] To ensure that the results within Group 2 were not due to lack of power, we randomly selected participants from the entire sample and created a subsample of 28 participants. This allowed us to test whether this sample size would be enough to detect an effect. Once again, we found a significant Group (1,2,3) x Remote (negative, positive) x Condition (attentive, inattentive) interaction, F (2, 25) = 3.5 p = .04, η_p^2 = .22, where children in Group 3 were significantly more likely to press the negative remote (M = 5.7, SE = 1.2) when no one was watching (p = .02), but were more likely to press the positive remote (M = 5.3, SE = 1.7) when the experimenter was watching (p = .07), albeit this was marginally significant. This provides further support that the three social-cognitive abilities collectively predict EAP.

perception (EAP). While a multitude of studies show EAP's effect on a wide range of behavior, we know surprisingly little about when and how this concern emerges in early development. In particular, when do children begin to show first signs of a concern for the potential evaluation of others? And what does it take to develop such a concern? The present dissertation explored potential social- cognitive abilities — including children's objectified sense of self, their theory of mind, and their norm understanding — that might predict the emergence of evaluative audience perception in 14- to 27-month-old children.

Using a within-subjects design, we tested children in Botto & Rochat's (2018) robot task as an index of evaluative audience perception, as well as six social-cognitive tasks that measured toddlers' objectified sense of self, theory of mind, and norm understanding. Based on a developmental framework that posits that the emergence of a new ability entails cognitive pre-requisites, we proposed that the combination of these three-specific social- cognitive abilities are potentially necessary for EAP to emerge. Specifically, we expected children who showed all three social-cognitive abilities to be more likely to strategically modify their button-pressing behavior based on the previously demonstrated values of the experimenter, as well as her relative attention (looking versus not looking at the child).

In support of this hypothesis, results showed that only children who displayed theory of mind, an objectified sense of self, and normativity also modified their button-pressing behavior depending on both the values and attention of the experimenter. In particular, we found that children who demonstrated all three cognitive abilities were more likely to press a remote that had been positively valued significantly more when the experimenter was watching, but then chose to press the remote that had been negatively valued significantly more once the experimenter had turned her back.

The specificity of the relation between the three abilities and EAP was further supported by the fact that other more general cognitive (i.e., spatial ability, language) and social factors (i.e., mother's level of self-consciousness) were not predictive of EAP. In particular, we found that performance on a spatial task, which we hypothesized should not be related to EAP, did not predict button-pressing behavior across conditions. This suggests that our results could not be explained by general intelligence or cognitive competency. Along the same vein, children's language ability, which is arguably a more general social-cognitive ability, was not related to EAP. Additionally, temperament and the parent's own self-consciousness were not related to button-pressing behavior, suggesting that parenting or the child's own traits (i.e., shyness or inhibitory control) might not be related to EAP. Lastly, as in previous studies (Botto & Rochat, 2018), we failed to find an age effect on button-pressing behavior across conditions of the robot task. We found this result to be surprising, as a 10-month span in late toddlerhood is large in terms of development. However, we interpret this finding, along with all the other null-findings in regard to other factors and EAP, as evidence that possessing all three proposed social-cognitive competencies is the most meaningful predictor of the emergence of EAP in toddlers.

Plausible Developmental Models

While this study provides support for the idea that the purported social-cognitive abilities are strongly related to the emergence of EAP, it does not speak to the potential causal role of these capacities, nor the directionality of the relation, as the study design is inherently correlational. Indeed, there are three potential developmental models that could be supported by these results (see Figure 4). One possibility, which is the one we favor, is that EAP is the result of children's development of theory of mind (Wellman, 2017), self-concept (Lewis, 1992), and normativity (Rakoczy & Schmidt, 2013). Such a model would suggest that the combination of

these three social-cognitive abilities give rise to EAP, and thus are necessary prerequisites to EAP. From this perspective, children's interindividual differences in theory of mind abilities, self-awareness, and sensitivity to norms might also influence the extent to which children care about others' evaluation, and thus express reputational concern (Chaplin & Norton, 2015). For example, if a child is better at understanding another's perspective or opinions (theory of mind), has a heightened self-awareness, and is sensitive to norms and standards, then they might be more sensitive to others' potential evaluations.

A second plausible developmental model is that seeing others as evaluators, coupled with an inclination to garner positive evaluations (i.e., EAP) could drive the development of an objectified sense of self, theory of mind and norm sensitivity. Indeed, if children are attuned to others' affective reactions, and are inclined to get positive affect from others, then they would be motivated to understand norms and the mental states of others. Accordingly, EAP would be important mechanism driving social development.

The third and last possible developmental model is that the three social-cognitive abilities and EAP simply co-emerge, suggesting that a more domain-general variable might underlie the development of an objectified sense of self, theory of mind, normativity, and EAP. The emergence of a secondary representation — which is the ability to go beyond one's immediate perceptual reality (i.e., what could be rather than what is) — could be such domain-general variable (Asendorpf, Werkenkin & Baudonniere, 1996; Perner, 1991). According to Perner (1991)'s influential model of representational development, toward the end of the second year, children develop the ability to hold two contrasting representations, and therefore can model hypothetical situations. This ability, which he referred to as a secondary representation, would allow children to go beyond representing the world with faithful accuracy and be able to engage

in pretend play, understand symbols, and construe others' minds. With this framework in mind, one could argue that for children to pass the mirror mark test, which is an index of an objectified sense of self, they must understand that the image reflected in the mirror (i.e., what is perceptually available), stands for themselves (secondary representation). Likewise, with theory of mind or normativity, children must also understand that someone's perceived actions represent unobservable desires, values or intentions (secondary representation).

While many theoretical accounts propose that secondary representations underlie imitation, theory of mind, and the development of a conceptual self (Meltzoff, 1990; Perner, 1991; Tomasello, 1995) there is little empirical support that a secondary representation is the driving mechanism behind their emergence and development. If a self-concept, theory of mind, normativity, and EAP, were driven by a secondary representation, then one would expect for these abilities to not only co-emerge, but also to be related, as they would presumably be sharing a common mechanism. Although there is some evidence that these social-cognitive competencies develop in tandem (Nielsen & Dissanayake, 2004), most studies have failed to find meaningful relations among these social-cognitive abilities (Agenetta & Rochat, 2004; Charman et al., 2000; Nielsen & Dissanayake, 2004). For example, Charman et al., (2000) investigated the longitudinal relation between imitation at 20 months and children's false belief understanding at 44 months and failed to find a significant link. Others have investigated the relation between mirror-self recognition and imitation and also found no relation (Agnetta & Rochat, 2004; Kristen-Antonow, Sodian, Perst, & Licata, 2015), although both of these cognitive competencies have been linked to pretend play (Nielsen & Dissanayake, 2004).

Moreover, results from the present study indicated that overall language competency, which would be considered a secondary representation as it is symbolic in nature, was not related

to an objectified sense of self, theory of mind, normativity, or EAP. Therefore, while a secondary representation might be an important component of the three social-cognitive correlates proposed here, there is little empirical support that a secondary representation is the driving factor for the emergence of these social-cognitive competencies and EAP.

Although the developmental models described above warrant further empirical investigation, evidence from individuals with autism lends some credence to the first developmental model (i.e., the three social-cognitive capacities are prerequisites to EAP). Children with autism, who are described as having a diminished sensitivity to others' evaluation, tend to not overimitate (see Edwards, 2014 for a review), and have deficits in theory of mind abilities (see Baron-Cohen, 2000 for a review). However, they do pass the mirror mark test (albeit, without signs of self-conscious emotions; Spiker & Rocks, 1984; Reddy, Williams, Costantini, & Lan, 2010). This dovetails with our proposal that the three social-cognitive capacities discussed here are necessary to display EAP, as we would expect for individuals with autism to also not develop EAP, because individuals with autism fail to manifest all three social-cognitive abilities. Indeed, there is little evidence to suggest that individuals with autism develop EAP (Baron-Cohen, 2010; Chevalier et al., 2012; Izuma et al., 2011).

Further, results from the present study showed that EAP was not predicted by each separate social-cognitive ability or particular combinations of two social-cognitive abilities. It was only children who manifested all three social-cognitive abilities that also displayed EAP as indexed by strategic-button pressing across conditions of the robot task. Such finding suggests that, on their own, each social-cognitive ability might be necessary, but not sufficient, to develop EAP. Notwithstanding, more research is needed to better understand the relation between theory of mind, normativity, a sense of self, and EAP (see below for Future Directions).

Implications

In addition to providing first evidence for the relation between EAP and particular social-cognitive abilities, these findings also have both theoretical and practical (translational) implications. From a theoretical standpoint, the findings speak to the possible nature of the cognitive process underlying EAP at its emergence. A lean interpretation might be that toddlers' strategic behavior in the robot task is rooted in eliciting positive rather than negative affect from others. While this interpretation would suggest that children are sensitive to, and distinguish between, positive and negative evaluations as we have defined them (i.e., sensitivity to another's response or reactions toward their behavior), it would not imply that children infer or attribute psychological states (i.e., desires or opinions) based on these affective responses. Alternatively, a richer interpretation would mean that by 24 months, children start to understand that others' emotional responses reflect a psychological state (as opposed to just a positive or negative affective response, such as a smile). That is, they can appreciate that positive reactions reflect a mental evaluation about the child. The fact that children who display EAP were also more likely to pass desires-based theory of mind tasks, which entails a representation of another's mental state (Repacholi & Gopnik, 1997; Wellman, 2017), supports this latter interpretation. Notwithstanding, future empirical investigations should probe the nature of EAP at its emergence, and how it develops to become more explicit.

A second theoretical implication is that the results of the current study put into question how early signs of self-conscious emotions in toddlers have been interpreted in the developmental literature. Until now, researchers considered early signs of embarrassment before the second year as a response to being the object of attention, presumably because toddlers did not have the cognitive capacities necessary to see others as evaluators (Lewis, 1992; Lewis &

Ramsay, 2002; Reddy, 2003). However, the current study challenges such interpretation, as we found that all of the putative precursors needed to develop evaluative audience perception — including an objectified sense of self, normative understanding and theory of mind — seem to emerge in the second year. Importantly, we found that the presence of all three of these social-cognitive abilities was related to the emergence of EAP, suggesting that sensitivity to others' evaluation is also evident in the second year. These results offer strong support for the idea that embarrassment, which is evident by the second year, might not be a simple stress response to being the object of attention as previously proposed, but could be considered as the expression of a sensitivity to how others might evaluate the child's behavior.

Beyond theoretical implications, the relation between the three social-cognitive abilities and EAP has the potential to inform our understanding of clinical disorders characterized by a diminished or heightened sensitivity to the evaluation of others, such as autism spectrum disorder and social anxiety disorder. Research shows that individuals with autism show a diminished concern for others' evaluations and fail to show reputational concerns in typical contexts (Chevalier et al., 2012; Izuma et al., 2011). For example, unlike typically developing adults and children, an observer does not affect autistic individuals' propensity to be generous (Izuma et al., 2011). In contrast, individuals with social anxiety manifest an exacerbated concern for others' potential evaluation (Rapee & Heimberg, 1997; Shenkler & Leary, 1982). Affecting more than 12% of the United States population, social anxiety is often debilitating in all areas of life, including relationships and careers, because the thought of being negatively evaluated induces high anxiety (Bögels, & Mansell, 2004; Kessler et al., 2005; Turk, Heimberg, & Hope, 2001).

Considering that these two disorders seem to illustrate extreme cases of evaluative audience perception (i.e., caring too little versus caring too much about another's evaluation),

understanding the emergence and cognitive correlates of evaluative audience perception in typically developing children might allow for earlier diagnosis and intervention for children deviating from the typical developmental trajectory. In terms of early diagnosis, measuring the extent to which a child displays evaluative audience perception in toddlerhood could serve as an early marker for later development of social anxiety (heightened sensitivity to others' evaluation) or autism spectrum disorder (diminished sensitivity to others' evaluation). In terms of developing possible interventions, the purported social-cognitive capacities related to EAP might explain major interindividual differences in the degree to which someone becomes sensitive to the evaluation of others. Thus, it could be possible that some symptoms of autism and social anxiety might be explained by deficits in theory of mind or normative understanding. In all, understanding the cognitive correlates of evaluative audience perception might shed light on the etiology of these clinical disorders, and provide some direction for possible interventions.

Limitations and Future Directions

Because this dissertation served as a first step to exploring the emergence and developmental underpinnings of evaluative audience perception, there are several limitations that should be addressed in future work. One such limitation is that some of the factors in this study, such as language competency, temperament, and the mother's self-consciousness were assessed via self-report questionnaires. While the measures we chose were empirically tested scales, it still begs the question as to whether direct observations might have yielded more accurate or different results. Future studies could asses self-consciousness and temperament dimensions using direct observations.

Future studies could also employ better tasks to measure an objectified sense of self and normativity. In particular, giving children a different version of the mirror mark test, or finding a

second normative task that might be more age-appropriate for toddlers, may better capture these cognitive abilities. In relation to measuring an objectified sense of self, we failed to find a relation between those who passed the mirror mark test and those who used personal pronouns. This could be due to the fact that the mirror mark test tends to have false negatives, as some children fail to reach for the mark, even if they do possess an objectified sense of self (Asendorpf, Warkentin & Baudonniere, 1996). One solution to this would be to use Asendorpf, Warkentin & Baudonniere (1996)'s revised version of the mirror mark test. After children looked at themselves in the mirror, the researchers also handed children a tissue and asked them to wipe themselves as an attempt to reduce the number of false negatives. With this revision, the number of passers increased from 45% to 61%. Therefore, this version might capture a more accurate representation of children's sense of self. For the normativity task, we chose to adapt a normativity task used by Clegg and Legare (2016) with older children to probe normativity. However, we found that only eight children passed this task, suggesting that the level of difficulty might be high for this age range. Future research might benefit from using a different normative task.

As we have previously mentioned, this study cannot provide evidence for a direct causal link between the three social-cognitive abilities and evaluative audience perception. Future studies could probe the plausibility of the three different models described here in several ways. One way is to investigate whether training children on the three social-cognitive abilities affect their display of evaluative audience perception. Ding et al. (2015) showed the efficacy of such training approach; when probing further the relation between theory of mind and children's ability to deceive, the researchers demonstrated that when children were trained on a theory of mind task, children's ability to lie improved. This served as strong evidence for the causal role

between theory of mind and lying. Along the same vein, researchers could train toddlers on theory of mind and normative understanding to see if such training has a direct effect on their display of EAP. A second avenue to explore the plausibility of the three developmental models, is to directly examine whether a secondary representation is indeed related to the emergence of all three social-cognitive abilities and EAP (i.e., the third developmental model). To do so, researchers could measure children's pretend play, an ability that has been shown to rely on secondary representations (Leslie, 1987; Perner, 1991; Suddendorf & Whiten, 2001), and investigate whether this ability predicts both EAP and the three social-cognitive competencies.

Lastly, the current dissertation only addresses when first implicit signs of evaluative audience perception emerge in development and its cognitive correlates. Specifically, an implicit evaluative audience perception entails an understanding of others as evaluators of the self, as well as a bias for positive as opposed to negative evaluations. However, the framework does not yet account for how this implicit form of evaluative audience perception becomes the more calculated and explicit evaluative audience perception geared towards reputational concerns seen in 4-5-year-olds. For example, an explicit evaluative audience perception would include children's ability to infer how other's evaluation might affect their future social interaction or another's concept of the child, as well as better understanding of societal norms or effective impression management strategies. Much work remains to comprehend this construct as it unfolds in development.

Conclusion

The concern for how others might evaluate our behavior is central to human psychology. Indeed, across contexts, humans deploy resources to manage their own image and reputation. However, until recently, little research had investigated the roots of such a concern. As a first

step to address this gap, this dissertation explored potential social-cognitive abilities that predicted the emergence of evaluative audience perception (EAP) in toddlers. Our results showed that, even after accounting for age, general cognitive competency, temperament and language, the combination of an objectified sense of self, normativity and theory of mind was the best predictor of EAP in 14- to 27-month-old toddlers. To our knowledge, this is the first study to provide evidence for the idea that these three social-cognitive competencies might be important for the emergence of EAP.

While this research serves as a critical first step to better understand the social-cognitive factors related to the emergence of EAP, many questions remain. In particular, what is the driving mechanism for the development of EAP in early development? And importantly, how does EAP evolve to become the more explicit reputational concern? As discussed here, exploring such questions could elucidate our understanding on the mechanism of EAP, as well as critical interindividual differences in the extent one becomes sensitive to the evaluation of others. Our concern with others' judgments is a central characteristic of human psychology. As such, it deserves much more empirical scrutiny.

References

Amsterdam, B. (1972). Mirror self-image reactions before age two. *Developmental Psychobiology, 5,* 297–305. http://dx.doi.org/10.1002/dev.420050403

Adedokun, O. A., & Burgess, W. D. (2012). Analysis of paired dichotomous data: A gentle introduction to the McNemar test in SPSS. *Journal of MultiDisciplinary Evaluation, 8*(17), 125-131.

Agnetta, B., & Rochat, P. (2004). Imitative games by 9-, 14-, and 18-month-old infants. *Infancy, 6*(1), 1-36.

Anderson, J. R., & Gallup, G. G. (2015). Mirror self-recognition: a review and critique of attempts to promote and engineer self-recognition in primates. *Primates, 56*(4), 317-326.

Asch, S. E. (1956). Studies of independence and conformity: A minority of one against a unanimous majority. *Psychological Monographs: General and Applied, 70,* 1-70. http://dx.doi.org/10.1037/h0093718

Asendorpf, J. B., Warkentin, V., & Baudonniere, P. M. (1996). Self-awareness and other-awareness. II: Mirror self-recognition, social contingency awareness, and synchronic imitation. *Developmental Psychology, 32*(2), 313.

Baillargeon, R., Scott, R. M., & Bian, L. (2016). Psychological reasoning in infancy. *Annual Review of Psychology, 67,* 159-186. https://doi.org/10.1146/annurev-psych-010213-115033

Bard, K. A., Todd, B. K., Bernier, C., Love, J., & Leavens, D. A. (2006). Self-awareness in human and chimpanzee infants: What is measured and what is meant by the mark and mirror test? *Infancy, 9*(2), 191-219.

Baron-Cohen, S. (2000). Theory of mind and autism: A fifteen-year review. *Understanding other minds: Perspectives from developmental cognitive neuroscience, 2,* 3-20.

Baumeister, R. F. (1982). Self-esteem, self-presentation, and future interaction: A dilemma of

reputation. *Journal of Personality*, *50*, 29-45. https://doi.org/10.1111/j.1467-6494.1982.tb00743.x

Bögels, S. M., & Mansell, W. (2004). Attention processes in the maintenance and treatment of social phobia: Hypervigilance, avoidance and self-focused attention. *Clinical Psychology Review, 24*(7), 827–856.

Botto, S. V., & Rochat, P. (2019). Evaluative Audience Perception (EAP): How Children Come to Care About Reputation. *Child Development Perspectives*, *13*(3), 180-185.

Botto, S. V., & Rochat, P. (2018). Sensitivity to the evaluation of others emerges by 24 months. *Developmental Psychology*, 54, 1723-1734, http://dx.doi.org/10.1037/dev0000548

Call, J., Carpenter, M., & Tomasello, M. (2005). Copying results and copying actions in the process of social learning: chimpanzees (Pan troglodytes) and human children (Homo sapiens). *Animal Cognition, 8(3), 151-163.*

Carlson, S. M., Davis, A. C., & Leach, J. G. (2005). Less is more: Executive function and symbolic representation in preschool children. *Psychological Science*, 16(8), 609-616.

Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen-through 18-month-old infants differentially imitate intentional and accidental actions. *Infant behavior and development*, *21*(2), 315-330.

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the society for research in child development, i-174.*

Chaplin, L. N., & Norton, M. I. (2015). Why we think we can't dance: Theory of mind and children's desire to perform. *Child Development*, *86*, 651-658. https://doi.org/10.1111/cdev.12314

Charman, T., Baron-Cohen, S., Swettenham, J., Baird, G., Cox, A., & Drew, A. (2000). Testing joint attention, imitation, and play as infancy precursors to language and theory of mind. *Cognitive development*, *15*(4), 481-498.

Chevallier, C., Parish-Morris, J., Tonge, N., Le, L., Miller, J., & Schultz, R. T. (2014). Susceptibility to the audience effect explains performance gap between children with and without autism in a theory of mind task. *Journal of Experimental Psychology: General*, *143*, 972-979. doi:10.1037/a0035483

Clay, Z., & Tennie, C. (2018). Is overimitation a uniquely human phenomenon? Insights from human children as compared to bonobos. *Child Development*, *89*(5), 1535-1544.

Clegg, J.M., & Legare, C.H. (2016). Instrumental and conventional interpretations of behavior are associated with distinct outcomes in early childhood. *Child Development*, 87(2), 527-542. doi: 10.1111/cdev.12472

Colonnesi, C., Rieffe, C., Koops, W., & Perucchini, P. (2008). Precursors of a theory of mind: A longitudinal study. *British Journal of Developmental Psychology*, *26*(4), 561-577.

Cordonier, L., Nettles, T., & Rochat, P. (2018). Strong and strategic conformity understanding by 3- and 5-year-old children. *British Journal of Developmental Psychology*. *36*, 438-451. https://doi.org/10.1111/bjdp.12229

Dadds, M. R., & Salmon, K. (2003). Punishment insensitivity and parenting: Temperament and learning as interacting risks for antisocial behavior. *Clinical Child and Family Psychology Review*, *6*, 69-86. Retrieved April 20, 2019 from http://www.ncbi.nlm.nih.gov/pubmed/12836578

Damon, W. (1983). Five questions for research in social- cognitive development. In E. T. Higgins, D. N. Ruble, & W. W. Hartup (Eds.), *Social cognition and social development: A socio-cultural perspective* (pp. 371–393). New York, NY: Cambridge University Press.

Dana, J., Weber, R., & Kuang, J. (2007). Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness. *Economic Theory, 33,* 67-80. https://dx.doi.org/10.1007/s00199-006-0153-z

D'Entremont, B., & Seamans, E. (2007). Do infants need social cognition to act socially? An alternative look at infant pointing. *Child Development, 78*(3), 723-728.

Ding XP, Wellman HM, Wang Y, Fu G, Lee K (2015). Theory-of-Mind Training Causes Honest Young Children to Lie. *Psychol Sci*. 2015;26(11):1812–1821. doi:10.1177/0956797615604628

Edwards, L. A. (2014). A meta-analysis of imitation abilities in individuals with autism spectrum disorders. *Autism Research, 7*(3), 363–380. doi:10.1002/aur.1379.

Engelmann, J. M., & Rapp, D. J. (2018). The influence of reputational concerns on children's prosociality. *Current Opinion in Psychology, 20, 92-95.*

Engelmann, J. M., Hermann, E., & Tomasello, M. (2012). Five-year olds, but not chimpanzees, attempt to manage their reputations. *PLoS One, 7*. e48433 https://doi.org/10.1371/journal.pone.0048433

Engelmann, J. M., Herrmann, E., & Tomasello, M. (2018). Concern for group reputation increases prosociality in young children. *Psychological Science, 29,* 181-190. https://doi.org/10.1177/0956797617733830

Evans, A. D., & Lee, K. (2013). Emergence of lying in very young children. *Developmental psychology*, *49*(10), 1958.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, *39*(2), 175-191.

Fenson, L. (2007). *MacArthur-Bates communicative development inventories*. Baltimore, MD:

Paul H. Brookes Publishing Company.

Fields, A. (2009). *Discovering statistics using SPSS*. 3rd edition. Sage, Los Angeles, CA

Flavell, J. H. (1999). Cognitive development: Children's knowledge about the mind. *Annual Review of Psychology*, 50, 21-45.

Friard, O., & Gamba, M. (2016). BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. *Methods in Ecology and Evolution*, *7*(11), 1325-1330.

Fu, G., Heyman, G. D., Qian, M., Guo, T., & Lee, K. (2016). Young children with a positive reputation to maintain are less likely to cheat. *Developmental Science*, *19*, 275-283. https://doi.org/10.1111/desc.12304

Gallup, G. G. (1970). Chimpanzees: Self-recognition. *Science*, 167, 86–87

George, D. and Mallery, P. (2010). *SPSS for Windows Step by Step: A Simple Guide and Reference 17.0* Update 10th Edition, Pearson, Boston.

Goffman, E. (1959). *The presentation of self in everyday life*. New York, NY: Random House.

Goulart, M., da Costa, N. C., Andrade, E. B., & Santos, A. A. (2015). Hedging against embarrassment. *Journal of Economic Behavior & Organization*, *116*, 310-318. https://doi.org/10.1016/j.jebo.2015.04.014

Haun, D., & Tomasello, M. (2011). Conformity to peer pressure in preschool children. *Child Development, 82*, 1759-1767. https://doi.org/10.1111/j.1467-8624.2011.01666.x

Heyman, G. D., Fu, G., & Lee, K. (2008). Reasoning about the disclosure of success and failure to friends among children in the United States and China. *Developmental Psychology*, *44*, 908-218. doi: 10.1037/0012-1649.44.4.908

Hilbrink, E. E., Sakkalou, E., Ellis-Davies, K., Fowler, N. C., & Gattis, M. (2013). Selective and

faithful imitation at 12 and 15 months. *Developmental Science*, *16*(6), 828-840.

Hoehl, S., Keupp, S., Schleihauf, H., McGuigan, N., Buttelmann, D., & Whiten, A. (2019). 'Over-imitation': A review and appraisal of a decade of research. *Developmental Review*, *51*, 90-108.

Huttenlocher, J., Newcombe, N., & Sandberg, E. H. (1994). The coding of spatial location in young children. *Cognitive Psychology*, *27*(2), 115-147.

Izuma, K., Matsumoto, K., Camerer, C. F., & Adolphs, R. (2011). Insensitivity to social reputation in autism. *Proceedings of the National Academy of Sciences*, *108*(42), 17302-17307.

Jones, E. E. (1964). *Ingratiation: A social psychological analysis*. New York, NY: Appleton-Century-Crofts.

Kagan, J. (1984). The idea of emotion in human development. In C. Izard, J. Kagan, & R. Zajonc (Eds.), *Emotions, cognition, and behavior* (pp. 38-72). Cambridge: Cambridge University

Kagan, J. (1981). *The second year: The emergence of self-awareness*. Harvard University Press.

Kagan, J., Reznick, J. S., Clarke, C., Snidman, N., & Garcia-Coll, C. (1984). Behavioral inhibition to the unfamiliar. *Child Development*, 2212-2225.

Kelsey, C. M., Grossmann, T., & Vaish, A. (2018). Early reputation management: Three-year-old children are more generous following exposure to eyes. *Frontiers in Psychology*, *9*, 698. https://doi.org/10.3389/fpsyg.2018.00698

Kenward, B. (2012). Over-imitating preschoolers believe unnecessary actions are normative and enforce their performance by a third party. *Journal of experimental child psychology*, *112*(2), 195-207.

Kenward, B., Karlsson, M., & Persson, J. (2011). Over-imitation is better explained by norm

learning than by distorted causal learning. *Proceedings of the Royal Society B: Biological Sciences, 278*(1709), 1239-1246.

Kessler, R. C., Berglund, P. D., Demler, O., Olga, J. R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*, 593–602.

Kristen-Antonow, S., Sodian, B., Perst, H., & Licata, M. (2015). A longitudinal study of the emerging self from 9 months to the age of 4 years. *Frontiers in psychology*, *6*, 789. https://doi.org/10.3389/fpsyg.2015.00789

Leary, M. R., & Allen, A. B. (2011). Belonging motivation: Establishing, maintaining, and repairing relational value. In D. Dunning (Ed.), *Frontiers of social psychology. Social motivation* (pp. 37-55). New York, NY: Psychology Press. https://doi.org/10.4324/9780203833995

Leimgruber, K. L., Shaw, A., Santos, L. R., & Olson, K. R. (2012). Young children are more generous when others are aware of their actions. *PLoS ONE, 7*, e48292. http://dx.doi.org/10.1371/journal.pone.0048292

Leslie, A. M. (1987). Pretense and representation: The origins of "theory of mind." *Psychological review*, *94*(4), 412.

Lewis, M (1992); The self in self-conscious emotions [Commentary on "Self evaluation in young children" by Stipek, D., Recchia, S., McClintic, S., & Lewis, M. *Monographs of the society for research in child development*, doi:10.1111/j.1540-5834.1992.tb00297.x

Lewis, M., & Ramsay, D. (2004). Development of self-recognition, personal pronoun use, and pretend play during the 2nd year. *Child Development*, *75*, 1821-1831. https://doi.org/10.1111/j.1467-8624.2004.00819.x

Lewis, M., & Ramsay, D. (2002). Cortisol response to embarrassment and shame. *Child Development*, *73*(4), 1034-1045.

Lewis, M., Sullivan, M.W., Stanger, C., & Weiss, M. (1989). Self-development and self-conscious emotions. *Child Development 60*, 146-156. http://dx.doi.org/10.2307/1131080

Lewis M., Brooks-Gunn J. (1979). *Social Cognition and the Acquisition of Self*. New York, NY: Plenum Press

Meltzoff, A. N. (2002). Imitation as a mechanism of social cognition: Origins of empathy, theory of mind, and the representation of action. *Blackwell handbook of childhood cognitive development*, 6-25.

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18 month old children. *Developmental Psychology*, 31, 1-16

Meltzoff, A. N. (1990). Foundations for developing a concept of self: The role of imitation in relating self to other and the value of social mirroring, social modeling, and self practice in infancy. In D. Cicchetti & M. Beeghly (Eds.), *The self in transition: Infancy to childhood* (pp. 139-164). Chicago: Chicago University Press

Mitchell, R. W. (1997). Kinesthetic-visual matching and the self-concept as explanations of mirror-self-recognition. *Journal for the Theory of Social Behavior*, 27(1), 17–39.

Moeller, M. P., & Schick, B. (2006). Relations between maternal input and theory of mind understanding in deaf children. *Child Development*, *77*(3), 751-766.

Newton, E. K., Laible, D., Carlo, G., Steele, J. S., & McGinley, M. (2014). Do sensitive parents foster kind children, or vice versa? Bidirectional influences between children's prosocial behavior and parental sensitivity. *Developmental Psychology*, *50*(6), 1808.

Nielsen, M., & Blank, C. (2011). Imitation in young children: When who gets copied is more important than what gets copied. *Developmental psychology*, *47*(4), 1050.

Nielsen, M., & Dissanayake, C. (2004). Pretend play, mirror self-recognition and imitation: A longitudinal investigation through the second year. *Infant Behavior and Development*, *27*(3), 342-365.

Over, H., & Carpenter, M. (2012). Putting the social into social learning: explaining both selectivity and fidelity in children's copying behavior. *Journal of Comparative Psychology, 126(2), 182.*

Over, H., & Carpenter, M. (2009). Eighteen-month-old infants show increased helping following priming with affiliation. *Psychological Science, 20,* 1189-1193. http://dx.doi.org/10.1111/j.1467-9280.2009.02419.x

Perner, J. (1991). *Understanding the Representational Mind*. Cambridge, MA: MIT Press.

Piazza, J., Bering, J.M., & Ingram, G. (2011). 'Princess Alice is watching you': children's belief in an invisible person inhibits cheating. *Journal of Experimental Child Psychology, 109,* 311–320.

Plotnik, J. M., De Waal, F. B., & Reiss, D. (2006). Self-recognition in an Asian elephant. *Proceedings of the National Academy of Sciences*, *103*(45), 17053-17057.

Putnam, S. P., Gartstein, M. A., & Rothbart, M. K. (2006). Measurement of fine-grained aspects of toddler temperament: The Early Childhood Behavior Questionnaire. *Infant behavior and development*, *29*(3), 386-401.

Rakoczy, H., & Schmidt, M. H. (2013). The early ontogeny of social norms. *Child Development Perspectives*, *7*, 17-21. doi:10.1111/cdep.12010

Rapee, R. M., & Heimberg, R. G. (1997). A cognitive-behavioral model of anxiety in social phobia. *Behaviour research and therapy*, *35*(8), 741-756.

Reddy, V. (2003). On being the object of attention: implications for self–other consciousness. *Trends in Cognitive Sciences*, *7*(9), 397-402.

Reddy, V., Williams, E., Costantini, C., & Lan, B. (2010). Engaging with the self: Mirror behaviour in autism, Down syndrome and typical development. *Autism*, *14*(5), 531-546.

Repacholi, B. M., & Gopnik, A. (1997). Early reasoning about desires: evidence from 14-and 18-month-olds. *Developmental Psychology*, *33*(1), 12.

Repacholi, B. M., & Meltzoff, A. N. (2007). Emotional eavesdropping: Infants selectively respond to indirect emotional signals. *Child Development, 78*, 503–521. https://doi.org/10.1111/j.1467-8624.2007.01012.x

Rochat P. (2018). The ontogeny of human self-consciousness. *Current Directions in Psychological Science, 27*, 345-350. doi: 10.1177/0963721418760236

Rochat, P. (2013). The gaze of others. In M. Banaji & S. Gelman (Eds.), *Navigating the social world: What infants, children, and other species can teach us* (pp. 205-211). New York, NY: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199890712.003.0037

Rochat, P. (2003). Five levels of self-awareness as they unfold early in life. *Consciousness and cognition, 12(4), 717-731.*

Rochat, P., Broesch, T., & Jayne, K. (2012). Social awareness and early self-recognition. *Consciousness and Cognition, 21*, 1491-1497. https://doi.org/10.1016/j.concog.2012.04.007

Scheier, M. F., & Carver, C. S. (1985). The Self-Consciousness Scale: A revised version for use

with general populations. *Journal of Applied Social Psychology*, 15, 687-699.

Schlenker, B. R., & Leary, M. R. (1982). Social anxiety and self-presentation: A conceptualization model. *Psychological Bulletin*, *92*, 641-669. https://doi.org/10.1037/0033-2909.92.3.641

Silver, I.M., & Shaw, A. (2018). Pint-sized public relations: The development of reputation management. *Trends in Cognitve Sciences*, *22*, 277-279. https://doi.org/10.1016/j.tics.2018.01.006

Sodian, B., Licata, M., Kristen-Antonow, S., Paulus, M., Killen, M., & Woodward, A. (2016). Understanding of goals, beliefs, and desires predicts morally relevant theory of mind: A longitudinal investigation. *Child Development*, *87*(4), 1221-1232.

Spiker, D., & Ricks, M. (1984). Visual self-recognition in autistic children: Developmental relationships. *Child Development*, 55, 214–225.

Stipek, D. J., Gralinski, J. H., & Kopp, C. B. (1990). Self-concept development in the toddler years. *Developmental Psychology*, *26*, 972-977. https://doi.org/10.1037/0012-1649.26.6.972

Stipek, D., Recchia, S., McClintic, S., & Lewis, M. (1992). Self-evaluation in young children. *Monographs of the Society for Research in Child Development*, *57*, 1-95. https://dx.doi.org/10.2307/1166190

Suddendorf, T., & Butler, D. L. (2013). The nature of visual self-recognition. *Trends in Cognitive Sciences, 17(3), 121-127.*

Suddendorf, T., & Whiten, A. (2001). Mental evolution and development: Evidence for secondary representation in children, great apes, and other animals. *Psychological Bulletin*, *127*(5), 629–650. https://doi.org/10.1037/0033-2909.127.5.629

Talwar, V., & Lee, K. (2008). Social and cognitive correlates of children's lying behavior. *Child*

development, 79(4), 866-881.

Tangney, J. P., Miller, R. S., Flicker, L., & Barlow, D. H. (1996). Are shame, guilt, and embarrassment distinct emotions?. *Journal of personality and social psychology*, 70(6), 1256.

Termini, R. B., & Tressler, L. B. (2008). American beauty: An analytical view of the past and current effectiveness of cosmetic safety regulations and future direction. *Food and drug law journal,* 63(1).

Thoermer, C., Sodian, B., Vuori, M., Perst, H., & Kristen, S. (2012). Continuity from an implicit to explicit understanding of false belief from infancy to preschool age. *British Journal of Developmental Psychology, 30, 172 –187. https://doi.org/10.1111/j.2044-835X.2011. 0206.x*

Tomasello, M. (2018). How children come to understand false beliefs: A shared intentionality account. *Proceedings of the National Academy of Sciences*, *115*, 8491-8498. https://doi.org/10.1073/pnas.1804761115

Tomasello, M. (1995). Understanding the self as social agent. In P. Rochat (Ed.), *Advances in psychology*, *112. The self in infancy: Theory and research* (p. 449–460). North-Holland/Elsevier Science Publishers. https://doi.org/10.1016/S0166-4115(05)80024-X

Turk, C.L., Heimberg, R. G., & Hope, D. A. (2001). Social anxiety disorder. *Clinical handbook of psychological disorders: A step-by-step treatment manual*, *3*, 114-153.

Vaughan, A., Mundy, P., Block, J., Burnette, C., Delgado, C., Gomez, Y., ... & Pomares, Y. (2003). Child, caregiver, and temperament contributions to infant joint attention. *Infancy*, *4*(4), 603-616.

Watling, D., & Banerjee, R. (2007). Children's understanding of modesty in front of peer and

adult audiences. *Infant and Child Development: An International Journal of Research and Practice, 16,* 227-236. https://doi.org/10.1002/icd.450

Wellman, H. M. (2017). The development of theory of mind: Historical reflections. *Child Development Perspectives*, *11*, 207-214. https://doi.org/10.1111/cdep.12236

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child development*, *75*(2), 523-541.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13, 103-128.*

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition,* 69(1), 1-34.

Zhao, L., Heyman, G. D., Chen, L., & Lee, K. (2017). Telling young children they have a reputation for being smart promotes cheating. *Developmental Science*, *21*: e12585, http://dx.doi.org/10.1111/desc.12585

Zilhão, J., Angelucci, D. E., Badal-García, E., d'Errico, F., Daniel, F., Dayet, L., ... & Murcia-Mascarós, S. (2010). Symbolic use of marine shells and mineral pigments by Iberian Neandertals. *Proceedings of the National Academy of Sciences*, *107*(3), 1023-1028.

Table 1

Summary of Tasks and Cognitive Measures Administered During Testing Session

Social-Cognitive Ability	Tasks	Measure	Passing Criteria	Scoring	References
Objectified sense of self	*Mirror Mark Test*	Measures mirror self-recognition by documenting children's self-referential behavior when seeing their reflection after being previously marked by rouge.	Child touches mark after seeing their reflection	Passing at least one of the two tasks counted as manifesting an objectified sense of self	*Amsterdam, 1972; Lewis et al., 1989; Rochat Broesch & Jayne, 212*
	Use of Personal Pronouns	Measures children's use of personal or possessive words via (i.e., me, you, mine) via parental reports (MacArthur-Bates).	Child uses at least 2 personal pronouns		*Stipek, Gralinski & Kopp, 1990*
Theory of Mind	*Imperative Pointing*	Measures infant's understanding of another's gaze and pointing gesture toward an object.	Child gives object of interest to experimenter	Passing at least one of the two tasks counted as manifesting theory of mind	*Carpenter, Nagell & Tomasello; 1998 Colonessi et al., 2008*
	Diverse Desires	Measures infant's understanding of diverse desires by testing whether child will give experimenter their preferred food.	Child gives experimenter her preferred food.		*Repacholi & Gopnik, 1997; Wellman & Liu, 2004*
Normativity	*Tower Task*	Measures children's propensity to faithfully reproduce a stacking sequence demonstrated by experimenter	Child reproduces correct stacking sequence	Passing at least one of the two tasks counted as manifesting normative understanding	*Clegg & Legare, 2016; Stipek et al., 1992*
	Overimitation	Measures child's propensity to faithfully copy arbitrary action (i.e., banging kaleidoscope on table)	Bangs kaleidoscope before looking through it.		*Call, Carpenter, & Tomasello, 2005; Over & Carpenter, 2012*
Control Task	*Spatial Task*	Measures children's search accuracy of a previously hidden toy in a 3-dimensional space after a brief delay. Accuracy score = distance away from toys' actual hidden location.	Accuracy score = distance away from toys' actual hidden location.	Average accuracy score across all five trials	*Huttenlocker, Newcombe & Sandberg, 1994*

Note. All tasks were coded as pass or fail, with the exception of the Spatial Task. Those who passed each task received 1 point. For the child to count as manifesting a particular social-cognitive ability, they had to have a minimum of 1 point within each social-cognitive ability (i.e., they had to pass at least one task). See Coding for details and rationale.

Table 2

Descriptive Statistics of Robot Task 1 and 2

	Mean	SD	Min	Max	Skewness	Kurtosis
Robot Task 1						
Attentive-Negative	4.72	5.3	0	19	1	-0.21
Attentive-Positive	4.75	5.1	0	21	1.2	0.97
Robot Task 2						
Attentive-Negative	2.15	3.75	0	23	1.1	0.26
Attentive-Positive	2.31	2.6	0	9	1.5	1.7
Inattentive-Negative	2.73	2.9	0	15	1.6	0.4
Inattentive-Positive	2.27	2.53	0	9	1.2	0.44

Note. Descriptive statistics pertain to raw button presses within each condition of the robot task. Attentive versus inattentive conditions describe whether the experimenter was observing the child. Although each robot task only had two conditions, the four rows under Robot Task 2 distinguish button-pressing between the two remotes (positive and negative) within each condition (attentive vs. inattentive). Robot Task 1 only used one remote.

Table 3

Descriptive Statistics and Relations of Cognitive Measures

Cognitive Ability	Passed (%)	Age (*p*)	Gender (*p*)	Relation (*p*)
Objectified sense of self				
Mirror Mark Test	16 (27)	.001*	.98	0.13
Personal Pronouns	24 (41)	.29	.08	
Theory of Mind				
Imperative Pointing	43 (73)	< .001*	.49	0.04*
Desires Task	31 (53)	.87	.17	
Normativity				
Overimitation	30 (50)	.89	.61	.001*
Tower Task	8 (13)	.02*	.79	

Note. P-values within the Age and Gender columns correspond to binary logistic regressions testing age and gender as predictors of performance in each task. The p-value under the column labeled Relation indicates the relation between the two cognitive tasks within each social-cognitive ability (McNemar test, two-sided)

* = $p < .05$

Table 4

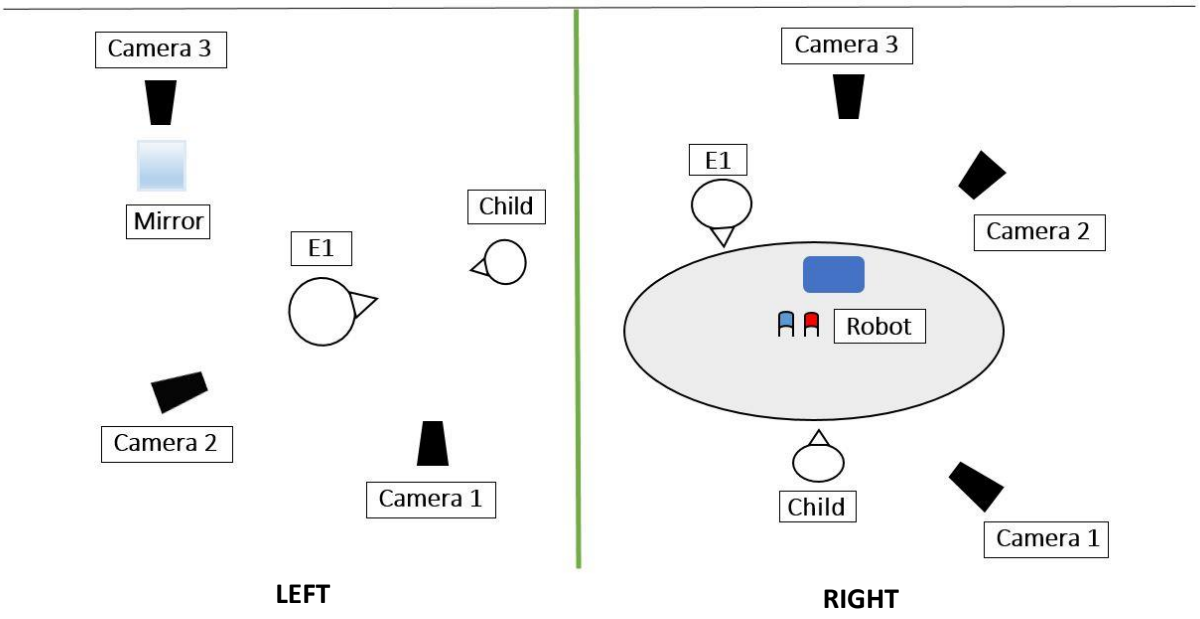
Descriptive Statistics for Questionnaires and Control Measures

	Mean	SD	Range	Min.	Max.	Skewness	Kurtosis
Temperament (ECBQ)							
Shyness	4.1	1.1	5.2	1.6	6.8	-0.2	-0.49
Inhibitory control	4	0.84	3.5	2.17	5.67	-0.07	-0.48
Sociability	5.4	0.94	4.5	2.5	7	-0.43	0.26
Self-Consciousness Scale							
Public	12.5	3.8	14	6	20	0.04	-1
Private	17.3	4.4	19	6	25	-0.62	0.12
Social Anxiety	8.2	4.8	16	0	16	0.01	-0.9
Spatial Score	7.1	3.2	13.43	1.13	14.56	0.38	-0.36
Language	30	27	82	0	82	0.52	-1.3

Note. Descriptive statistics reflect raw scores for each measure. Language scores reflect the percentage of vocabulary words the child pronounced and understood (see Coding for details).

Figure 1

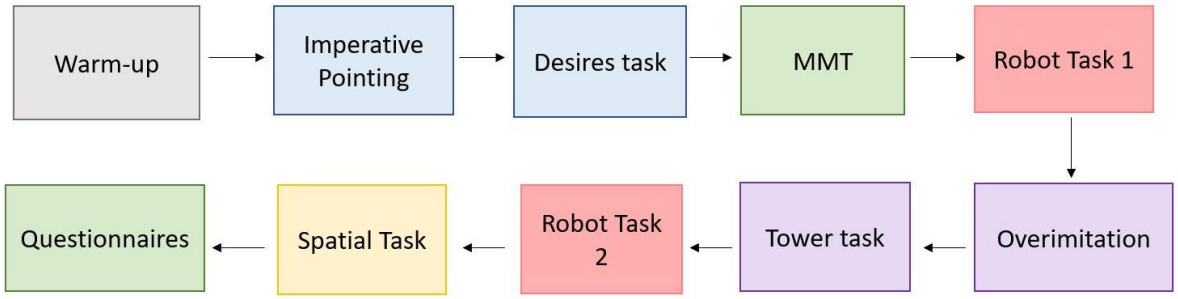
Bird's Eye-View of Experimental Room Set-Up



Note. The testing room was divided by a green curtain as pictured above. The first part of the study took place on the left side of the room. Tasks administered on the left side included imperative pointing, desires task, and mirror mark test. The second part of the study took place on the right side of the room. Tasks included both robot tasks, both normativity tasks, and the spatial task.

Figure 2

Order of Evaluative Audience Perception and Cognitive Tasks Administered in Testing Session



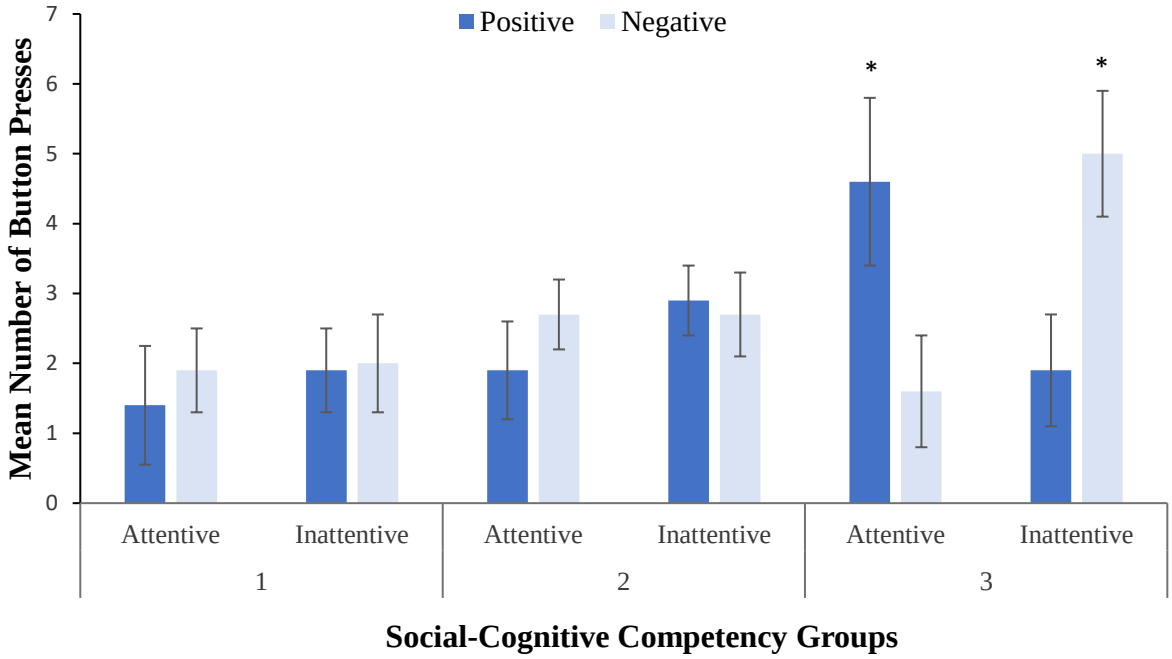
Legend

Red Evaluative Audience Perception
Blue Theory of Mind
Green Objectfied sense of self
Purple Normativity

Note. All participants received all of the tasks in the same order. Each individual block indicates a task, and each color indicates the cognitive ability that the task measured.

Figure 3

Button-Pressing Behavior in Conditions of Robot Task 2 Across Social-Cognitive Groups

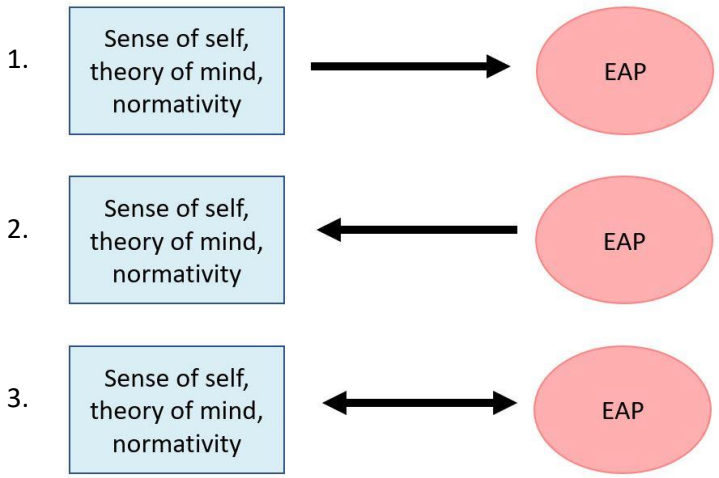


Note. Bars represent standard error. A mixed ANOVA revealed a significant Remote (positive, negative) x Condition (attentive, inattentive) x Cognitive Group (Group: 1, 2, 3) interaction, F (2, 56) = 5.9, p = .005, η_p^2 = .18, where children in Group 3 were the only group to display EAP as indexed by strategic button-pressing in the robot task.

* = p < .05 after Bonferroni corrections (i.e., adjusted p-values).

Figure 4

Potential Developmental Models of Evaluative Audience Perception



Note. Illustration of three plausible developmental models that could explain the relation between the three proposed social-cognitve abilities and EAP. The first two models illustrate a causal role between the cognitive abilities and EAP. In model 1, the three cognitive abilities would be pre-requisites to EAP. In model 2, EAP would be a driving mechanism behind the development of the three cognitve abilities. The last model (Model 3) illustrates a co-emergence between the three cognitive abilities and EAP, suggesting that a more domain-general variable might account for their emergence and development.