2023-02-13

Simulating Peers: Can puppets simulate peer interactions in studies on children's socio-cognitive development?

Stengelin, R

10.1111/cdev.13913
Child Development
Wiley

DOI: 10.1111/cdev.13913

REGISTERED REPORT

CHILD DEVELOPMENT

Simulating peers: Can puppets simulate peer interactions in studies on children's socio-cognitive development?

Roman Stengelin[1,2,3] | **Daniel B. M. Haun**[1,2] | **Patricia Kanngiesser**[4]

[1]Department of Comparative Cultural Psychology, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

[2]Leipzig Research Center for Early Child Development, Leipzig University, Leipzig, Germany

[3]Department of Psychology and Social Work, University of Namibia, Windhoek, Namibia

[4]School of Psychology, University of Plymouth, Plymouth, UK

Correspondence
Roman Stengelin, Department of Comparative Cultural Psychology, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany.
Email: roman_stengelin@eva.mpg.de

Funding information
Max-Planck-Gesellschaft. This research was funded by the Max Planck Society.

Abstract
Interactions with peers are fundamental to socio-cognitive development, but assessing peer interactions in standardized experiments is challenging. Therefore, researchers commonly utilize puppetry to simulate peers. This Registered Report investigated urban German children's (Age_{Range} = 3.5–4.5 years; N = 144; 76♀) mind ascriptions and social cognition to test whether they treat puppets like peers, adults, or neither. Children attributed less mind properties to puppets than peers or adults. However, children's social cognition (i.e., normativity, prosociality, and theory of mind) varied little across partners. Puppetry relies on children's ability for pretense, but can provide valid insights into socio-cognitive development. Implications for using puppets as stand-ins for peers in developmental research are discussed.

From early on in development, children need to learn how to navigate their social world. To do so proficiently, they rely on social-cognitive abilities that enable them to adhere to the norms of their cultural group, help those around them, and understand others' intentions and beliefs (Callaghan et al., 2011; Nielsen & Haun, 2016; Tomasello, 2019). Many of these abilities emerge during early to middle childhood and allow children to grow into competent members of their cultural groups. Children rely on interactions with partners, both adults and peers, to develop their social-cognitive skills. While the study of social interactions in adult-child settings allows for rigorous control of partners' behaviors, peer settings are more challenging to standardize and control. Therefore, researchers commonly utilize hand puppets, animated by adult experimenters, to simulate peers (*puppetry*). While this approach is common practice in child development research, experimental work testing whether children's behaviors toward puppets mimic those toward peers is still lacking. The current study aims to bridge this gap by investigating (i) whether children's social-cognitive behaviors (i.e., normativity, prosociality, and theory of mind) differ when interacting with puppets as compared to peer or adult partners, and (ii) if potential variation in social cognition may derive from differences in children's mind ascriptions to these agents.

Social interactions with peers and adults

Interactions with peers and adults provide young children with fundamental yet often complementary social experiences. During infancy and toddlerhood, adults provide children with support and advice to help them acquire new skills and knowledge (Csibra & Gergely, 2009; Kline, 2015). While adults continue to be

Abbreviations: AC, adult condition; CC, child condition; HPD, highest posterior density; PC, puppet condition; WAIC, Watanabe-Akaike information criterion.

Registered Report: (Stage 2, Stage 1 has been accepted in-principal/IPA at *Child Development*; see https://osf.io/dqycm).

essential partners throughout childhood, peer interactions become increasingly frequent and important drivers of social-cognitive development from age three to four (Brownell et al., 2006; Jaswal & Neely, 2006; Rakoczy et al., 2010; Tomasello & Gonzalez-Cabrera, 2017).

Children's social experience with peers differs from their experience with adults. There is an apparent gap between the expertise provided by adults and peers. Relationships with adults are usually hierarchical with predefined social roles and adults scaffolding children's learning. By contrast, peer interactions are more egalitarian, with children being sometimes more, sometimes less knowledgeable or competent than their partners (e.g., Brownell et al., 2006). It is important to note that this pattern derives from ethnocultural beliefs and parenting practices common in societies that emphasize formal education and direct pedagogy (Keller, 2007; Keller & Kärtner, 2013; Rogoff, 2003). Although interaction patterns of children growing up in other contexts may vary considerably (Hewlett, 1996; Lancy, 2008; Rogoff, 2016; Terashima & Hewlett, 2016), previous studies have stressed that across a range of cultural contexts, interactions with peers *and* adults provide significant experiences and learning opportunities throughout the preschool years (Broesch et al., 2021; Keller, 2007; Lewis et al., 1984; Lew-Levy et al., 2019; Rogoff, 2003).

Preschoolers' behaviors vary depending on whether they interact with peers or adults. They typically perceive adults as more knowledgeable than peers (Taylor et al., 1991) and, from toddlerhood, children often prioritize information provided by adults over information provided by peers (Jaswal & Neely, 2006; Kachel, Moore, et al., 2018; Molleman et al., 2019; Rakoczy et al., 2010; Zmyj et al., 2012). Three- and 4-year-old children also enforce rules learned from adults more often than rules learned from peers (Rakoczy et al., 2010). At the same time, preschoolers also start to show some flexibility in their social learning when, for example, peers are more reliable than adult informants or have more expertise in a domain (Jaswal & Neely, 2006; VanderBorght & Jaswal, 2009). Yet, when solving complex problems, children usually benefit from engaging with adults. For example, Rogoff and colleagues found that 5- to 9-year-olds showed more sophisticated collaboration and became more efficient in errand action planning tasks when paired with adults compared with peers (Gauvain & Rogoff, 1989; Radziszewska & Rogoff, 1988, 1991).

The egalitarian structure of preschoolers' peer interactions might elicit behaviors that would normally not (yet) be shown with adults. For example, children's protest toward peer transgressors (Köymen et al., 2014) may precede their protest toward adult transgressors (e.g., Heyman et al., 2016). Many studies on this topic assume that children are more reluctant to protest against adults (compared with peers) because they perceive them as authority figures who know and set the rules (Rakoczy et al., 2010; Rakoczy & Schmidt, 2013). Such perceived hierarchies may also shape other domains of social-cognitive development. In a recent study, 2- to 3-year-old German children helped peers at higher rates than adults (Ulber & Tomasello, 2020). The authors assumed that these children perceived their peers as less competent than adults, increasing their motivation to help (see also Hepach et al., 2017). Finally, there is tentative evidence that children are more likely to ascribe false beliefs to peers than adults due to assumed differences in general competence (Seehagen et al., 2018). In sum, these findings suggest that preschoolers' social cognition varies depending on whether they engage with adults or peers.

Methodological challenges in the study of peer interactions

Despite the importance of social interactions with peers and adults for preschooler's development, most experimental protocols in early childhood research focus on social interactions between children and adults. Often this approach is less guided by theoretical concerns but more by a focus on experimental rigor and internal validity: Put simply, researchers can train adult experimenters to strictly adhere to study procedures, allowing for standardized test situations. Ensuring such high levels of standardization becomes more challenging in experimental peer settings. Some studies have relied on (spontaneous) peer interactions (e.g., Kachel, Moore, et al., 2018; Kachel, Svetlova, et al., 2018; Rekers et al., 2011), but these approaches can only be used in a narrow range of test situations. Moreover, these studies often struggle to standardize behaviors of peer partners across participants and, therefore, need to analyze findings on a dyadic level, requiring larger sample sizes and rendering some interpretations difficult. One approach to standardizing peers' behaviors in experiments is to present participants with pre-recorded video stimuli of peers (van Leeuwen et al., 2018; Zmyj et al., 2012). However, children's performance often decreases when watching videos as compared to live interactions (Anderson & Pempek, 2005; Reiß et al., 2019), and video stimuli rarely allow for natural communication and interaction (Nielsen et al., 2008; but see Kachel et al., 2021). Another approach is to extensively train child confederates (e.g., Engelmann et al., 2016, 2018). However, such training is challenging and time-consuming as peer confederates require sustained practice, need to interrupt their daily routines to assist with the research endeavor, and are more prone to errors and deviations from the scripted procedures than adult experimenters (see also Zmyj et al., 2012).

Using puppetry to simulate peer interactions

To overcome the challenges of experimental peer studies, researchers often rely on puppets, animated by adults, as proxies for peers (e.g., Gampe & Daum, 2018; Kanngiesser & Warneken, 2012; Rakoczy et al., 2008; Stengelin et al., 2018; Warneken & Tomasello, 2013). One critical assumption of this approach is that puppetry reduces the asymmetry between children and adult experimenters, boosting children's self-confidence and eliciting behaviors and skills they would otherwise only show with peers (Benenson, 1993; Simon et al., 2008). Children are assumed to be "comfortable interacting with hand puppets and show realistic social behaviors" (Warneken & Tomasello, 2013, p. 343), implying an ideal trade-off between external and internal validity in these studies.

In studies on normativity, researchers often present participants with a puppet whose behavior deviates from established rules (e.g., Hardecker et al., 2016; Rakoczy et al., 2008, 2009; Schmidt et al., 2016; Vaish et al., 2011). In a seminal study, children learned how to play a novel game (Rakoczy et al., 2008). Next, a puppet (played by an adult) approached the scene and made a mistake according to the game's rules. Children protested against the puppet, indicating a normative interpretation of said rules. Similarly, children's prosociality has also been frequently assessed using puppetry (see also Dunfield et al., 2013; Green et al., 2018; Kim et al., 2014; Plötner et al., 2015). For example, Warneken and Tomasello (2013) investigated whether children would choose to instrumentally help a puppet fulfill its needs. In seminal studies on children's false belief understanding, researchers illustrated agents' mental states through puppets and dolls engaging in play-like activities (Baron-Cohen et al., 1985). More recent work in this domain continues to use puppets to simulate social interactions in a child-friendly manner (see Grueneisen et al., 2017; Moll et al., 2017; Rubio-Fernández & Geurts, 2013). Interestingly, a meta-analysis by Wellman et al. (2001) indicated no systematic variation in young children's false belief reasoning regarding real people or puppets.

Puppets as animate and mindful agents?

Using puppetry to simulate peer interactions forms part of a widespread practice of using puppets, dolls, or animated objects to represent social agents in experimental child research (see Kominsky et al., 2022 for a recent discussion). This practice taps into children's tendency for anthropomorphism, the bias to ascribe human-like features to non-human entities (i.e., robots, see Damiano & Dumouchel, 2018; or animals, see Urquiza-Haas & Kotrschal, 2015). For example, Dolgin and Behrend (1984) found that preschoolers ascribed animacy to dolls and nonhuman animals (i.e., other mammals and birds) alike, but that older children understood that dolls lacked animacy. The authors emphasized that children may rely on physical properties and self-movement to attribute animacy to objects (Jipson & Gelman, 2007). Furthermore, a recent study found that 4-year-olds actively demonstrated actions to a puppet only if an adult experimenter had previously treated the puppet as an agent rather than as an object (Asaba et al., 2022). This study indicates not only that children conceive of puppets as animate agents (see also Kominsky et al., 2022), but also that they treat them as mindful partners under appropriate circumstances. It is thus likely that children would also view puppets as possessing mental attributes like human agents: A cognitive capacity to think (*agency*) and a more emotional and motivational capacity to feel (*experience*; Gray et al., 2007; Moriguchi et al., 2019; Weisman et al., 2018).

While puppetry offers a powerful tool to ensure experimental rigor in empirical research, claims of external validity—the degree to which children's behaviors toward puppets resemble those toward real-world social partners—remain mostly untested. Although researchers have argued that puppetry elicits children's motivation and attention (Epstein et al., 2008; Rudolph & Heller, 1997; Simon et al., 2008), such notions have mostly relied on anecdotal evidence and experiences of researchers, practitioners, and educators. In a qualitative study, Epstein et al. (2008) assessed the usage of hand puppets to elicit children's talk in clinical settings. Children were given a choice to talk to an adult interviewer, a hand puppet enacted by an adult, or to animate a hand puppet themselves. Children commonly preferred to communicate with puppets rather than adults. While this research indicates that children like to engage with puppets (see also Dorie et al., 2013; Remer & Tzuriel, 2015; Simon et al., 2008), it does not provide a quantitative assessment of whether children treat puppets like real-life social partners.

In sum, puppetry offers an approach for simulating non-hierarchical (peer) interactions, which is commonly used and widely accepted among developmental researchers. For the most part, claims that children's behaviors toward puppets resemble real-world social interactions reference previous work utilizing the same approach (Grueneisen et al., 2017; Huber et al., 2019; Stengelin et al., 2018; Warneken & Tomasello, 2013). Only recently have researchers started to question this assumption (i.e., Heyman et al., 2016; Revencu & Csibra, 2020). By the end of their second year of life, children already understand pretense behaviors (Walker-Andrews & Kahana-Kalman, 1999) and may, therefore, conceive puppets as peers if introduced accordingly (Sutherland & Friedman, 2012; Thompson & Goldstein, 2020). However, a mature understanding of pretense should also enable them to discern that the agent animating the puppet causes the puppet's behaviors. If this was the case, one would expect them to interact similarly with a puppet and an adult. Alternatively, children may conceive of puppets as distinct entities that neither resemble peers nor adults. For example, children may probe behaviors they would not show otherwise because they are engaged in playful and fun activities with an adult.

To our knowledge, no study has directly addressed these alternatives. In the present study, we undertook such as quantitative evaluation by assessing children's behaviors when engaging with puppets, adults, and peers by focusing on three essential domains of social cognition: Normativity, prosociality, and theory of mind.

The current study

We utilized a mixed design to investigate whether children's social-cognitive skills and behaviors with puppet partners resemble those with peers, adults, or neither. As a potential mechanism of children's social cognition across partners, we furthermore assessed the extent to which children ascribe mental properties to puppets compared with adults and peers. We evaluated three hypotheses denoting that children treat puppets as peers (henceforth: puppet-as-child hypothesis), adults (puppet-as-adult hypothesis), or as distinct agents (puppet-as-puppet hypothesis). Research hypotheses and study protocols were peer-reviewed and published as a Registered Report (Stengelin et al., 2021).

Accordingly, this study seeks to validate the use of puppets in experimental childhood research. As most previous work in these domains has utilized puppets to study child development in the Global North (Nielsen et al., 2017), the study was conducted with children from an urban German context in which children experience an abundance of puppets and other toys in their everyday life. It is beyond the scope of the current study to test if and how puppetry research can be generalized outside such cultural contexts.

The study was divided into two phases: In the *mind ascription phase*, we assessed children's attributions of mental properties to puppets, peers, and adults. We focused on preschoolers' evaluations of agents' cognitive (*agency*) and emotional (*experience*) capacities as previous research has shown that preschoolers view them as the main properties of human minds (Gray et al., 2007; Moriguchi et al., 2019; Weisman et al., 2018). In the subsequent *behavioral phase*, children were randomly assigned to one of the following conditions (between-subjects factor): (1) in the puppet condition (PC), children interacted with a puppet animated by an adult experimenter; (2) in the child condition (CC), children interacted with a peer; (3) in the adult condition (AC), children interacted with an adult experimenter. We focused on three domains of social-cognitive development, assessed within subjects: Normativity, prosociality, and theory of mind.

Contrasting participants' behaviors in the CC and the AC allowed us to conceptually replicate and extend previous research on the effects of partner's age on children's socio-cognitive abilities (i.e., Seehagen et al., 2018; Ulber & Tomasello, 2020). More importantly, relating the PC to both the CC and AC allowed us to understand how puppetry can be generalized to social interactions with real partners. As a proxy for children's normativity, we assessed their protest behavior when partners violated an established rule (Rakoczy et al., 2008, 2010). As a proxy for prosociality, we analyzed children's tendency to instrumentally help their partners (Ulber & Tomasello, 2020; Warneken & Tomasello, 2013). Finally, children's false belief understanding served as a proxy for their theory of mind (Baron-Cohen et al., 1985; Seehagen et al., 2018).

We formulated the following confirmatory hypotheses regarding children's social cognition: We expected children to show higher protest rates toward peers than toward adults (protest: CC > AC; see Heyman et al., 2016; Rakoczy et al., 2008). Moreover, we expected them to help peers at higher rates than adults (instrumental helping: CC > AC; see Ulber & Tomasello, 2020). Third, we predicted that children would ascribe false beliefs to peers more often than adults (false belief: CC > AC; see Seehagen et al., 2018).

To validate puppetry as an approach to simulate peer interactions, we further predicted (confirmatory hypotheses) that children's behaviors in the PC would resemble those displayed in the CC. Applying this puppet-as-child hypothesis to children's social cognition, we expected them to show similar levels of protest (normativity), instrumental help (prosociality), and false belief understanding (theory of mind) in the PC and the CC (PC = CC).

The contrasting puppet-as-adult hypothesis held that children conceive of puppets as a proxy for adults who animate the puppet. Accordingly, we assumed children's protest (normativity), instrumental helping (prosociality), and false belief reasoning (theory of mind) to be at similar rates for adult partners and hand puppets (PC = AC).

As a third alternative, the puppet-as-puppet hypothesis proposed that neither of these hypotheses would apply. Following this hypothesis, we expected to find behaviors in the PC at intermediate levels compared with the CC and the AC (CC > PC > AC). Alternatively, puppets may mark playful contexts with adults in which children probe behaviors without facing negative consequences, such as retaliation (Heyman et al., 2016). If so, children would show some behaviors at higher rates in the PC than in the AC and the CC (PC > CC ≠ AC).

We also applied these hypotheses when investigating children's mind ascriptions to puppets, peers, and adults. Here, we expected that the majority of children would ascribe mental properties in both the *agency* and *experience* dimension to human partners (adults and peers; see also Moriguchi et al., 2019), but that they might be reluctant to also attribute these properties to puppets (PC < CC = AC). Combining children's mind ascriptions and their social-cognitive behaviors in a single study would allow us to draw firm conclusions about the ecological validity of puppetry in child development research. If, for example, children would ascribe similar mind properties to puppets and peers and, at the same time, show similar behaviors to both of these partners, this would provide strong support for puppetry as a valid approach to simulate peers. Support for the puppet-as-puppet hypothesis regarding both children's mind ascriptions and their socio-cognitive behaviors would, in contrast, raise doubt on the ecological validity of puppetry as an approach to simulate social interactions.

While we did not expect children to treat puppets as peers in only some of the social-cognitive domains tested here, we were aware that our study might support the puppet-as-child hypothesis for some but not all domains tested. For example, children might accept puppets as peer-like partners in normative contexts, given that such contexts create a playful setting in which puppets alleviate children's initiative. Simultaneously, children might be more reserved to help a puppet partner as the situation may not be perceived as a playful setting.

METHOD

Participants

We tested a total of N = 149 (M_{Age} (SD) = 4.12 (0.35); 76♀) children to gain a minimum of n = 48 datapoints per behavioral task for each of the three between-subject conditions (PC, CC, AC). A further n = 13 children (4 per condition and 1 additional child in the PC to train a new experimenter) were tested during a piloting phase to ensure experimenter training and fine-tuning of the study protocol. We recruited children of a narrow age range of 3.5 to 4.5 years, given the increased importance of peer interactions at this age and the assumed variation in children's social cognition within this age range. Data were collected between October 2021 and May 2022.

Testing took place in a child laboratory in Leipzig, a mid-sized German city with a population of approximately 593,000 people. Children were recruited from a database of parents who had given their consent to be contacted for child development studies. We invited an additional $n_{confederates}$ = 49 children as partners (matched by gender and age to their peer partners) in the CC. We randomly determined for each dyad which child would be the focal test child/partner. As anticipated, children came from mid- to high socio-economic backgrounds and more than 80% of parents reported to have graduated from high school. More than 90% of the city's residents are German citizens, which also applies to the vast majority of children in the database.

Testing appointments were arranged via phone calls. In case two families agreed to participate on the same date, children were tested in the CC after ruling out that both participants knew each other. In case one of the participants needed to cancel their appointment, or if only one participant could be scheduled for a given date, children were randomly assigned to either the PC or AC. We followed this recruitment strategy until we reached the total sample size in each condition. To be included in the data analyses, children needed to complete the initial mind ascription phase and at least two out of three behavioral tasks. Reasons for drop-outs included children's unwillingness to participate and experimenter errors. We excluded children from the normativity task if their partner did not sort according to their rule at test or if the test child realized that two different rules exist (as evidenced by their utterances). In the prosociality task, we excluded children if their partner did not require help (i.e., because they found a way to unlock the box themselves, see procedure) or if test children did not learn how to open the box themselves (see procedure). Furthermore, technical issues (e.g., camera not working and malfunction of study apparatuses) were treated as potential issues leading to data exclusion. We documented each drop-out (see Supporting Information) and substituted the participants until we reached the full sample size (N_{min} = 48 for each task). Please note that this approach led to minimally larger sample sizes in some tasks.

Ethical statement

The study procedure fell under the approval of the internal ethics review board at the Department of Comparative Cultural Psychology at the Max Planck Institute for Evolutionary Anthropology; no reference number assigned. The study protocol strictly adhered to the legal requirements of psychological research in Germany. Parents provided written consent before their child participated in the study, and children were be asked for their assent. Children received small gifts for their participation.

Study design and materials

Our mixed-design study consisted of two study phases. In the mind ascription phase, we assessed children's mind ascriptions to unfamiliar puppets, peers, and adults (*condition*, assessed within subjects). We used a set of six laminated DIN A5 portraits depicting a puppet, a peer, and an adult (each photograph matched for children's gender). All depicted agents were unfamiliar to children (i.e., a different puppet was depicted than the one used as children's partner in behavioral tasks), and we adopted questions used in previous studies on children's mind ascriptions to different agents (Gray et al., 2007; Moriguchi et al., 2019; see below).

In the behavioral phase, we combined a between-subjects factor *condition* (PC, CC, and AC) with a within-subjects factor *task* (normativity, prosociality, theory of mind) to test children's behaviors when engaging with these partners. Here, we counterbalanced the order of tasks across participants. Overall, the study lasted about 30 to 45 min per session. In the normativity task, the test child sorted wooden blocks by placing them on two vertical sticks attached to a wooden board. We used 16 wooden blocks varying in shape and color (four red balls, four blue balls, four red cubes, and four blue cubes) based on previous studies (Kanngiesser et al., 2022; Köymen et al., 2014). To reduce potential noise (and improve sound quality for data coding) when placing the blocks on the sticks, we attached small foam pieces to each block's top and bottom. Furthermore, the test child received two transparent plastic containers with an equal number of blocks of each shape and color.

In the prosociality task (see also Ulber & Tomasello, 2020), we used a wooden box (15 × 15 × 5 cm) together with six marbles. The box's top could be locked and opened using a metal key. Furthermore, we used a golden paper mâché toy elephant with a xylophone inside (as used in Stengelin et al., 2020). By dropping marbles into the trunk of the elephant, the xylophone produced jingling sounds.

For the theory of mind task, we tested children's false belief understanding by using a candy box (e.g., "Smarties"; see Gopnik & Astington, 1988) filled with five wooden pencils as the unexpected content.

We used a *kumquat Living Puppet®* (size = 65 cm) with blonde hair as children's partner in the PC. Such puppets are commonly utilized in childhood research to simulate peer interactions (Grueneisen et al., 2017; Huber et al., 2019; Stengelin et al., 2018; Warneken & Tomasello, 2013). Experimenters could simultaneously insert one hand at the back of the puppet's head to animate the puppet's face (head direction and mouth) while using another hand to animate the puppet's arm and hand.

We covered the other tasks' materials with a blanket to prevent distraction during each task. Both partners (i.e., the test child and the peer/puppet/adult) received balloons after each behavioral task and all participating children received a small gift at the end of the study (see Figure 1 for illustrations of the partners in each condition and the test situations across behavioral tasks).

Procedure

An adult experimenter (E1) of the opposite gender as the test child conducted the study. We matched test children's study partners for gender. A second experimenter (E2) of the test child's gender operated the puppet in the respective condition. In the AC, E2 acted as the partner. In the CC, a same-gender peer served as the partner.

Warm-up phase

Upon arrival in the laboratory, E1 welcomed each child and their parent(s) in a reception room. E1 explained the study to the parent(s) and asked for their written consent. Furthermore, E1 warmed up the test child and familiarized them with the laboratory setting by asking some questions (e.g., "What is your name?", "What did you do today?") and by playing with a standardized set of toys. Next, E1 introduced the test child to their partner. In the CC, E1 introduced both test child and partner by their names and asked them about their ages, emphasizing that both were of similar age; in the PC and AC, E1 introduced the partner as "Alex" (introduced as he or she to match the test child's gender). In the PC, the puppet was animated by E2 throughout the study and treated as a child by E1. E1 asked the test child for their name, and the puppet stated that "I am your age." In the AC, E1 also asked the test child for their age. In this condition, E2 emphasized being "much older than you; I am an adult." After the test child or partners showed clear signs of familiarization (talking to E1 and moving freely in the warm-up area), E1 presented a short game to both partners involving a small ball and a bucket. E1 explained that "in this game, we will try to throw the ball into the bucket. But we must stand behind this line when throwing the ball. It is not so easy. You can try it yourselves!". This task was sufficiently challenging for 3- to 4-year-old children such that they would miss the bucket most of the times. Thus, children experienced that they and others were not infallible (see also Rakoczy et al., 2008). Each partner had the opportunity to throw the ball two times or until either partner had missed the bucket at least once. Once completed, E1 praised children for playing well and asked if they wanted to engage in more activities. E1 asked both partners to wait for more tasks to be set up. Children then participated in the mind ascription phase before engaging in three behavioral tasks (presented in a counterbalanced order):

Mind ascription phase

Next, the test child entered the testing room and sat down on a cushion. E briefed the child that they would like to hear what they thought of some people, presenting three photographs of unfamiliar same-gender agents (i.e., a puppet, an adult, and a peer). Next, E1 put two of these photographs aside, referring to the remaining agent (order of presentation was randomized across participants). For the peer photograph, E1 said "Look, do you see this one?" before asking the child on their mind ascriptions regarding the peer. It has to be noted that the German translation of this utterance conveys information on the peer's gender, which was matched to the child: "Schau mal, siehst du den/die da?". To assess children's ascription of *experience*, E asked "Can this one feel hungry?" and "Can this one feel pleasant?". To assess ascriptions of *agency*, E asked "Can this one remember?" and "Can this one communicate, such as talking with someone?". These questions have indicated good factor load in the original

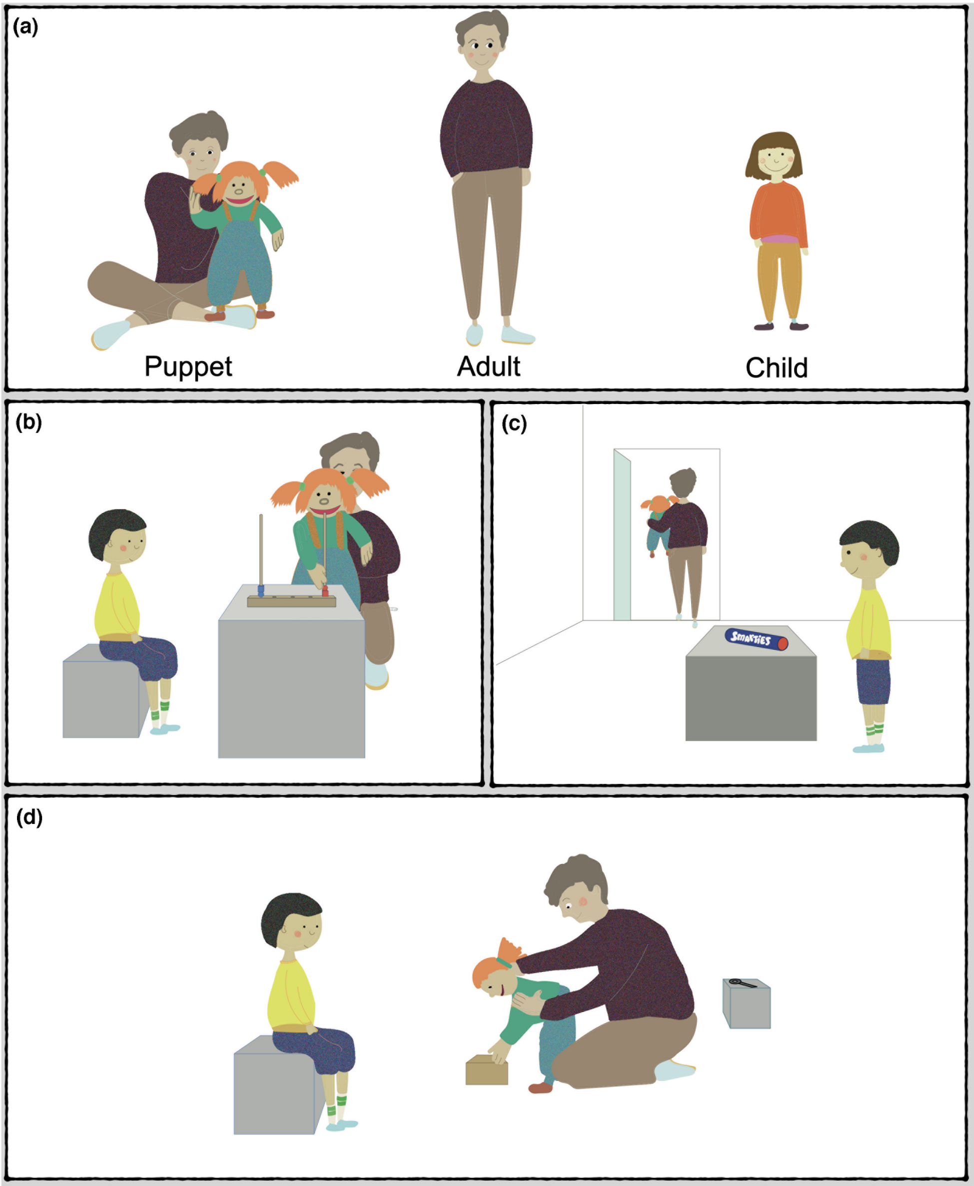

FIGURE 1 Schematic images of study design and experimental set-up (behavioral tasks); (a) partners across conditions (partner's gender will be matched to test child's gender); (b) set-up in normativity task; (c) set-up in theory of mind task; (d) set-up in prosociality task; set-ups in (b–d) resemble the test situations in the puppet condition; the test child is depicted wearing a yellow shirt.

study by Gray et al. (2007) and have proven efficient to assess children's mind ascriptions regarding each dimension among preschool-aged children (Moriguchi et al., 2019). Identical instructions were used to introduce adult and puppet stimuli. Mind dimensions were introduced in a random order (see also Moriguchi et al., 2019). This study phase lasted around 5 min.

Behavioral phase—Normativity task

In a training session, the test child entered the testing room while their partner waited in the reception room. The test child sat on a chair in front of the sorting game (the study protocol of this task was adopted from Kanngiesser et al., 2022; Köymen et al., 2014). Each vertical stick displayed two wooden blocks (i.e., a red ball and a blue cube), and E1 modeled sorting the blocks according to color. After sorting two blocks (i.e., same-colored ball and cube) on each side, E1 asked the test child to sort the remaining 10 blocks in the presence of E1. If the test child failed to do so, E1 repeated the instruction once more. Children only proceeded if they sort correctly. Finally, E1 removed all blocks except the last one on each stick, and the child sorted the blocks by themselves. Once the test child completed their sorting, they were asked to recall the sorting rule before returning to their parents in the reception room, and the partner learned the sorting game. During this time, E1 ensured that the test child did not brief their partner on the game's sorting rule.

In the CC, the peer partner's instructions were identical, except that the partner learned to sort the blocks according to shape. Again, the partner saw two blocks (a red ball and a blue cube) placed on the sticks. After observing E1 sorting two blocks on either side (i.e., red cube and ball vs. blue cube and ball), partners were given turn to sort the remaining 10 blocks themselves under E1's supervision. Next, E1 removed newly sorted blocks from the sticks. Partners were asked to recall their sorting rule. In the two remaining conditions (PC and AC), the puppet or adult partner waited some minutes inside the room.

Next, the test child returned to the testing room while their partner waited on a chair next to the sorting game. Before entering, E1 reminded them of their respective sorting rule (i.e., color). The test child sat down on a table approximately 1.5 m distance from the sorting game. E1 instructed the test child to observe the partner and the partner to sort "according to the rule." Next, E1 left the room for 90s. In this period, we assessed children's protest behavior as the dependent variable of this task. After E1's return, they debriefed both partners on the two different sorting rules and thanked both players for their participation while handing them one balloon. This task lasted around 10 min.

Behavioral phase—Prosociality task

In this task, the test child and their partner entered the testing room together with E1 to play a "marble game" (the task procedure of this task was based on the instrumental helping task in Ulber & Tomasello, 2020, with minor modifications to match the anticipated age range and the overall study protocol). A locked wooden box was placed on the floor and a golden toy elephant located on an adjacent table in the room, with four marbles next to it. E1 encouraged both the test child and their partner to put the marbles into the elephant's trunk (i.e., "feeding" the elephant). After each partner had inserted two marbles, the test child was asked to remain in the test room, while the partner was invited to join parents in the reception room.

Next, E1 rattled the wooden box in front of the test child. E1 told the test child that another marble was locked in the box, which could be used to "feed the hungry elephant." Next, E1 showed the child that the box could be unlocked with a metal key placed on a chair in one corner of the room. E1 explained how to open the box and modeled the opening actions. After this, E1 encouraged the child to "feed the elephant" with the just-received marble from the other box. In the meantime, E1 placed another marble in the box and locked it. Handing the key to the child, E1 encouraged the child to open the box herself to "feed the elephant." This procedure was repeated once more, with the key being placed on the adjacent chair. E1 offered help if the child failed to open the boxes (e.g., "What do you need to open the box?"; "Where do you find the key?"; see Ulber & Tomasello, 2020). After opening the box without external support for two consecutive trials, children were asked to draw a picture to avoid ceiling effects in children's helping behaviors. Simultaneously, E1 locked another marble in the box and hid the key in the previous location. E1 further stated that "this marble is meant for [name of partner]. You can draw in the meantime!".

E1 asked the partner child to return to the test room and rattled the box to indicate that it contained a marble. E1 stated that "I need to leave briefly, but you [partner] can open the box and feed the elephant in the meantime." No further instruction was given, and E1 returned to the room after 1 min. If the partner had no success in unlocking the box, E1 would help them to "feed the elephant." E1 thanked both partners for their participation, handing them one balloon each. This task took around 8 min.

Behavioral phase—Theory of mind task

To assess children's false belief understanding, we ran an unexpected content false belief task, the "Smarties" task (Gopnik & Astington, 1988; see also Wellman & Liu, 2004). E1 invited the test child into the testing room and showed them a "Smarties" box. Next, E1 asked them what they thought was inside the candy box. If the child did not reply, E1 prompted them, asking: "Does it look like there would be candy inside?". As a second prompt, E1 asked the child, "what kind of box is this? What could be in here?". Finally, E1 could ask the child: "Could there be candy or pens in here?". E1 would only proceed with the task if the test child assumed there were candy or "Smarties" inside the box. Next, E1 showed the child the candy box's content and, acting surprised, announced that "there are some pens inside!" (showing the pens to the child). E1 then closed the candy box and asked the child to recall the content of the box. If the child failed to respond correctly, E1 would show the contents again and assess the child's recall. E1 noted that "[the partner] has never looked inside the candy box" before proceeding with the test questions: "What does [the partner] think is in the box? Candy or Pens?". E1 further asked the child whether their partner had looked inside the candy box as a memory check. E1 then showed both partners the candy box and asked what they thought was inside the box. Regardless of their responses, both children received a balloon each. The task took up to 5 min.

Post-test phase

Following the three behavioral tasks, E1 thanked participants and their parents for their participation and debriefed children about the study. Parents could ask questions about the study. Children received a small gift that they could choose from a set of small toys (e.g., books, toys, and balls).

Coding

A research assistant, blind to hypotheses and not otherwise involved in the study, coded the entire sample. A second research assistant, also blind to hypotheses, coded 25% of the preregistered sample size (n = 36; six boys and six girls per condition) for reliability purposes. The first author trained raters on a subset of six participants from the piloting phase (two per condition). Raters were instructed to approach the first author in case of unclear cases by describing the situation and transcriptions without referring to the condition to ensure unbiased feedback. As an indicator of coding reliability, we calculated Cohen's kappa (κ) for all (dichotomous) outcomes. In case of disagreement between raters, we used the first research assistant's coding for our data analyses.

Mind ascription phase

For each agent and question, we coded participants' affirmative responses regarding each mind dimension (*experience*, *agency*; see Gray et al., 2007) as 1 and negative responses as 0. For example, a child indicating that a puppet can "remember" but cannot feel "hungry" will be given a score of 1 in the *experience* domain for the first response and a 0 in the *agency* domain (later response). As such, each child received a total of four dichotomous scores per agent, indicating their mind ascription (two per dimension). Interrater reliability was excellent (all κ's ≥ .98).

Behavioral phase—Normativity task

We coded participants' protest from the moment the partner touched the first block to E1's return to the testing room. In line with previous coding schemes (Köymen et al., 2014), we scored children's utterances as *protest* if they intervened in either a normative way (coded verbatim, such as using normative vocabulary, e.g., "It does not go like this") or in an imperative way (such as instructing the partner, e.g., "Not like this!", "Do it like this!"; detailed data on normative and imperative protest are provided in the Supporting Information and OSF). Gestures and physical corrections of the partner's behaviors were not coded as *protest*. We scored the presence of *protest* as 1 (irrespective of its frequency) and absence as 0. Interrater reliability was perfect (κ = 1).

Behavioral phase—Prosociality task

We scored test children's *instrumental helping* as 1 if they had helped their partner unlocking the box during E1's absence. This included physical help (e.g., getting and handing over the key) or verbal help (e.g., pointing to the key or referring to it verbally). If children showed no attempt to help, they received a score of 0. Interrater reliability was excellent (κ = .94).

Behavioral phase—Theory of mind task

We scored test children's *false belief understanding* as 1 if they stated that their partner would think that the candy box contained candy/"Smarties" and while remembering the box's real content (i.e., pens). Failure to respond to both questions correctly was scored as 0. Interrater reliability was perfect (κ = 1).

Data analyses

We performed all statistical analyses within a Bayesian framework in *R* (R Core Team, 2018) using the *brms* package (Bürkner, 2017, 2018). All reported methods and analyses were preregistered unless specified otherwise. Thus, the current approach can be considered confirmatory. To test the degree to which children's behaviors in the three social cognition tasks varied across conditions, as well as to which extent their mind ascriptions varied across agents, we ran generalized linear mixed models with Bernoulli response distributions using probit link functions. We fitted separate models for each task using the binary outcome (mind ascription: experience and agency; behaviors: protest, instrumental helping, false belief reasoning) using the function *brm*.

To investigate children's mind ascriptions, we fitted models comprising the predictor condition (PC vs. CC vs. AC) while controlling for children's age (in years and standardized) and gender (coded as −0.5/girls and 0.5/boys to allow for better interpretation of the estimates). We added children's ID as a random intercept to the model and fit a random slope of condition per item question (see procedure above). We also preregistered to add a random slope of condition within each mind dimension (agency vs. experience). However, due to small cell sizes per mind dimension and item question, this model could not be fit reliably. We thus decided to remove this random slope from our main analysis. To test potential effects of mind dimensions on children's mind ascriptions, we conducted a separate exploratory analysis in which we fit an interaction of the fixed effects of condition and mind dimension. Since this analysis indicated no effect of dimension on children's mind ascriptions (see below), further analyses did not focus on this predictor. We estimated separate intercepts for each level of condition (rather than dummy—code the predictor) to ensure that effects could be estimated for each partner. To illustrate the effects of the predictors and their potential interactions, we report the estimates of the model output, the 89% highest posterior density (HPD) interval from these models and plot the posterior distribution (McElreath, 2020).

For each of the behavioral tasks, we fitted a model comprising the predictor condition (PC vs. CC vs. AC) as well as the control variables age (in years and standardized) and gender (coded as −0.5/girls and 0.5/boys to allow for better interpretation of the estimates). We added task order as a random intercept to the model. We estimated separate intercepts per condition to focus on each level of this predictor. As in the previous analysis, we illustrate the effects of condition and the control variables on the outcomes by reporting the estimates of the model output, the 89% HPD interval from these models, and plotting the posterior distribution (McElreath, 2020). Additionally, we report the posterior probability that the intercept estimated for the PC is closer to the intercept of the CC as compared to the intercept of the AC.

Furthermore, we directly assessed whether children's mind ascriptions and behaviors affirmed the puppet-as-child hypothesis, the puppet-as-adult hypothesis, the puppet-as-puppet hypothesis, or whether their behaviors were robust to their partner's characteristics (i.e., PC = AC = CC). We took two approaches to this. First, we tested whether the above models' 89% HPD intervals for the difference in the probability of the behavior of interest (i.e., protest, instrumental helping, and false belief reasoning) between the PC and CC excluded 0. If so, this would be interpreted as evidence *against* the puppet-as-child hypothesis. The results of this test are straightforward to interpret, but like all such tests prone to false positive/negative results due to sampling variation (see "Power analysis" below). Secondly, we ran separate models for each domain and hypothesis with models comprising similar predictors and controls as the model described above but with *condition* recoded accordingly. That is, the models comprising information on all three conditions (PC vs. CC vs. AC) represent the puppet-as-puppet hypothesis (henceforth: $Model_{puppet\text{-}as\text{-}puppet}$). To model the puppet-as-child hypothesis, we ran a model contrasting children's behavior in the CC/PC and their behavior in the AC (i.e., $Model_{puppet\text{-}as\text{-}child}$). Furthermore, we contrasted the CC and AC/PC ($Model_{puppet\text{-}as\text{-}adult}$) to represent the puppet-as-adult hypothesis. Finally, we fitted a $Model_{no\ condition}$ lacking variation regarding condition (CC/AC/PC) to represent the assumption that children's behaviors were indifferent to partners. Comparing these models allowed us to directly assess the relative strength of evidence for each of our hypotheses provided by the data.

We used the widely applicable information criterion/Watanabe-Akaike information criterion (WAIC) scores and weights to describe and relate these models (see also Bohn et al., 2020; for a similar approach). Smaller WAIC scores indicate better out-of-sample predictive accuracy and thus better fit. We used the WAIC weights of the full models to assess which of the four models ($Model_{puppet\text{-}as\text{-}puppet}$, $Model_{puppet\text{-}as\text{-}child}$, $Model_{puppet\text{-}as\text{-}child}$, $Model_{no\ condition}$) best predicted new data (all WAIC weights add up to 1). It has to be noted that this approach allowed us to present a gradual estimation of the equivalence between conditions, rather than dichotomous support for either statistical equivalence or difference regarding each outcome.

Power analyses

To estimate the power of the first approach to correctly detect differences between PC and the CC (contradicting the Puppet-as-Child hypothesis), we ran simulation-based power analyses in which we assessed the sensitivity of our study design to detect varying underlying probabilities (i.e., effect sizes) between the two conditions. Doing so, we estimated up to which "true" effect size separating the PC and the CC (i.e., varying probabilities to show the behavior of interest) the 89% HPD intervals would exclude 0. We simulated 1000 data sets, each comprising $n = 48$ subjects per condition (behavioral phase) or $n = 144$ per condition (mind ascription phase), and varied the probabilities of the behavior of interest in each condition. We estimated the probabilities in the AC and CC based on findings from previous studies using similar approaches and conditions (Köymen et al., 2014; Moriguchi et al., 2019; Seehagen et al., 2018; Ulber & Tomasello, 2020) and accounted for uncertainty in these estimates. It has to be noted that the probabilities used to simulate these effects, including that of condition, would most likely diverge from the actual data we obtained. Reasons for this may have included publication bias in previous studies or diverging procedures between previous investigations and the current study. To compensate for this, we modeled the uncertainty of these effects within our power analyses by sampling values from a normal distribution centered on values described in previous work.

We further simulated participants' age according to the anticipated age range (3.5–4.5 years), scaled this variable, and simulated a positive effect of age on the behavioral outcomes following previous work reporting age-related increases in each domain (Köymen et al., 2014; Seehagen et al., 2018; Ulber & Tomasello, 2020). Thus, we applied a uniform distribution to limit the space of potential age effects between probabilities of 0 and 0.1. As we counterbalanced gender and task order (for behavioral tasks), these variables did not need to be simulated but were set instead. To allow for potential effects of gender, we applied a normal distribution setting this effect within the probability range of −.1 to .1. To allow for potential effects of task order on the behavioral outcomes, we simulated such effects by, for each simulated dataset, first sampling a standard deviation from an exponential distribution with an expected value of 0.1 and then sampling three values from a normal distribution with mean zero and the sampled SD. These three values were then centered on zero (to preserve the interpretation of per-condition intercepts as overall means), resulting in typically weak effects of task order, with half of simulated datasets having per-order deviations from the overall mean response probability of less than about .03.

It is important to note that findings in which the 89% HPD intervals would include 0 regardless of a simulated "true" effect would not lead us to assume practical equivalence between conditions. Instead, such findings would be analyzed based on model comparisons, yielding gradual support for either practical equivalence between the PC and the CC or systematic variation across these conditions.

Mind ascription phase

Following data reported by Moriguchi et al. (2019), we set the probabilities for children's ascription of mind dimensions to adults at $p_{\text{Experience AC}}$ = .72 and $p_{\text{Agency AC}}$ = .75. For children, we set probabilities of $p_{\text{Experience CC}}$ = .73 and $p_{\text{Agency CC}}$ = .62 (aggregated based on ascriptions to babies and adults). We set similar probabilities following the puppet-as-child hypothesis ($p_{\text{Experience PC}}$ = .73; $p_{\text{Agency PC}}$ = .62) and added uncertainties of SD = 0.10 for all simulations to account for procedural differences across studies (i.e., peer partners, rather than babies; gender-matched partners). Probabilities ranging from −.3 to .3 (not exceeding 1) were added as effect sizes to simulate systematic differences between the CC and the PC.

Based on 1000 simulations, the power analysis suggested that a sample size of n = 144 had a power of 0.80 to reveal an effect of 0.11 between the CC and the PC (see Supporting Information). For smaller effect sizes, the 89% HPD interval may include 0. Model comparisons would reveal gradual support for practical equivalence or systematic variation across conditions.

Behavioral phase—Normativity task

We simulated the probability of a subject to protest in the AC as $p_{\text{Protest AC}}$ = .35 based on previous research on children's protest against adult transgressors (Heyman et al., 2016). Given the procedural difference in this study and the current study, we let this probability vary with an SD = 0.10 to reflect uncertainty regarding this probability. The probability of protesting in the CC was set at $p_{\text{Protest CC}}$ = .67 based on previous work assessing children's protest toward peers (Köymen et al., 2014). Given the procedural similarity between the original and the current study, we let this probability vary with an SD = 0.05. It should be noted that the estimated difference between these probabilities for the AC and the PC is likely conservative given that the two studies differed in the severity of transgressions (Heyman et al., 2016: moral transgression; Köymen et al., 2014; conventional transgressions). Thus, one might have expected to find more considerable differences between the AC and the PC in the current study.

The mean probability of protest in the PC was set according to the puppet-as-child hypothesis at $p_{\text{Protest PC}}$ = .67. We added effect sizes reflecting probabilities between −.3 and .3 (but ensured probabilities could not exceed 1) to simulate probabilities for the PC in order to assess whether our analyses would be sensitive to systematic differences between the CC and the PC.

Based on 1000 simulations, the power analysis suggests that a sample size of n = 48 per condition has a power of 0.80 to reveal an effect size of 0.22 between the CC and the PC (see visualization of power analyses in the Supporting Information). That is, for true effects smaller than 0.22, the 89% HPD interval might include 0. In this case, model comparisons would reveal which hypothesis would be best explained by the data.

Behavioral phase—Helping task

We simulated the probability of a test child to help their partner in the AC as $p_{\text{Help AC}}$ = .42 based on previous research using a similar approach (Ulber & Tomasello, 2020). For the CC, we set the probability at $p_{\text{Help CC}}$ = .75 based on the same research. Given the similarity of the current procedure with that of previous research, we added an uncertainty of SD = 0.05 to simulate data for both the AC and the CC. Again, we set the mean probability of test children to helping in the PC at $p_{\text{Help PC}}$ = .75 according to the puppet-as-child hypothesis and added probabilities between −.3 and .3 (ensuring that probabilities could not exceed 1) as effect sizes to vary between the CC and the PC.

Power analyses following 1000 simulations suggested that studies with a sample size of n = 48 per condition would detect an effect size of 0.21 (i.e., 89% HPD interval excluding 0) with a power of 0.80 (see Supporting Information). For smaller effect sizes, the 89% HPD interval might include 0, and model comparisons would reveal gradual support for practical equivalence or systematic variation across conditions.

Behavioral phase—Theory of mind task

Based on the study by Seehagen et al. (2018), we set the probability of test children's false belief reasoning with adult partners at $p_{\text{False Belief AC}}$ = .27. Based on the same study, we set a probability of $p_{\text{False Belief CC}}$ = .43 for the CC. We added uncertainties of SD = 0.10 for both simulations to account for differences between previous work and the current study (i.e., sum scores aggregated for different false belief tasks). The probability for the PC was set accordingly to the puppet-as-child hypothesis ($p_{\text{False Belief PC}}$ = .43). Probabilities ranging from −.3 to .3 (not exceeding 1) were added as effect sizes to simulate systematic differences between the CC and the PC.

Based on the 1000 simulations, the power analysis suggested that a sample size of $n = 48$ per condition had a power of 0.80 to reveal an effect between the CC and the PC for an effect size of 0.24 (see Supporting Information). For smaller effect sizes, the 89% HPD interval may include 0. Model comparisons would reveal gradual support for practical equivalence or systematic variation across conditions.

RESULTS

Mind ascription phase

Children ascribed some mind properties to all partners, but they were more likely to attribute agency and experience to humans, both peers and adults, than puppets (see Table 1 for proportions of mind ascriptions per partner and mind dimensions).

We did not find any indication that children's mind ascriptions to puppets, peers, and adults varied across the dimensions of agency and experience ($Fixed\ Effect_{Mind\ Dimension}$: Estimate ± SE = −0.03 ± 0.09; $Fixed\ Effect_{Condition Child \times Mind Dimension}$: Estimate ± SE = −0.01 ± 0.10; $Fixed\ Effect_{Condition\ Puppet \times Mind\ Dimension}$: Estimate ± SE = −0.02 ± 0.09; see also Table 1; Figure 2). Thus, the following results refer to statistical models predicting children's mind ascriptions on an item level, lacking information about these dimensions.

Children were more likely to attribute mind properties to adults ($Fixed\ Effect_{Adult}$: Estimate ± SE = 1.35 ± 0.36) and peers ($Fixed\ Effect_{Child}$: Estimate ± SE = 1.14 ± 1.26) than to puppets ($Fixed\ Effect_{Puppet}$: Estimate ± SE = 0.23 ± 0.40). The posterior probability that children's mind ascriptions to puppets were more similar to their responses to children than to their responses to adults was 0.58.

Further analyses revealed strong evidence for the puppet-as-puppet hypothesis regarding children's mind ascriptions (see Figure 2). First, the model's 89% HPD interval for the difference in the probability of mind ascriptions between the PC and the CC very nearly excluded 0 (89%-HPD = [−0.52; 0.00]). In a similar vein, WAIC-comparisons showed much better predictive accuracy of the $Model_{puppet\text{-}as\text{-}puppet}$ (WAIC ± SE = 1520.1 ± 45.7; weight > .99) compared with the $Model_{puppet\text{-}as\text{-}child}$ (WAIC ± SE = 1664.8 ± 43.6; weight < .01), the $Model_{puppet\text{-}as\text{-}adult}$ (WAIC ± SE = 1673.8 ± 43.1; weight < .01), and the $Model_{no\ condition}$ (WAIC ± SE = 1721.5 ± 42.9; weight < .01). Taken together, these results provide clear evidence that children ascribe less agency and experience to puppets compared with peers or adults.

TABLE 1 Mean proportions of mind ascription across partners and mind dimensions.

	Mind dimension	
Partner	Agency	Experience
Puppet	.51	.60
Adult	.84	.82
Child	.79	.84

Behavioral phase

Normativity task

Children protested against puppets (prob. = .80), peers (prob. = .89), and adults (prob. = .82) at high rates. Model estimates suggested no obvious effect of condition on children's protest behavior ($Fixed\ Effect_{Puppet}$: Estimate ± SE = 0.73 ± 0.33; $Fixed\ Effect_{Child}$: Estimate ± SE = 1.15 ± 0.35; $Fixed\ Effect_{Adult}$: Estimate ± SE = 0.84 ± 0.32). The posterior probability that children would protest against puppets as against children rather than adults was .25. Thus, it was unlikely that children consider puppets as fully-fledged stand-ins for peers in the domain of normativity.

Further analyses confirmed this impression (see Figure 3a). The model's 89% HPD interval for the difference in protest probability for the PC and CC just included zero (89%-HPD = [−0.25; 0.03]). WAIC-comparisons indicated that the most parsimonious $Model_{no\ condition}$ (WAIC ± SE = 123.9 ± 14.5; weight = .37) had slightly better predictive accuracy than the $Model_{puppet\text{-}as\text{-}adult}$ (WAIC ± SE = 124.2 ± 14.6; weight = .31). Both models outperformed the $Model_{puppet\text{-}as\text{-}puppet}$ (WAIC ± SE = 125.7 ± 14.7; weight = .15) and the $Model_{puppet\text{-}as\text{-}child}$ (WAIC ± SE = 125.5 ± 14.3; weight = .17). It is worth noting that the differences in WAICs across these models were markedly smaller than the models' WAIC's standard errors, indicating rather subtle variation in predictive accuracy across models. Children's protest with puppets did not differ from their protest with adults, as assumed by both the $Model_{no\ condition}$ and the $Model_{puppet\text{-}as\text{-}adult}$. However, there remained some uncertainty as to how children's protest in peer settings relates to their protest with puppets and adults. That is, children might protest slightly more in peer settings than when engaging with other partners. However, such an effect would be markedly small, demanding for (exceptionally) large sample sizes in experimental child psychology research to be detected reliably. Given the advantage in predictive accuracy and the parsimony of the $Model_{no\ condition}$, we conclude that children's protest with puppets matches their protest with peers and adult partners.

Helping task

Children helped puppets (prob. = .92), peers (prob. = .81), and adult partners (prob. = .88) at high rates. There was no obvious effect of condition on children's helping behavior ($Fixed\ Effect_{Puppet}$: Estimate ± SE = 1.32 ± 0.31;

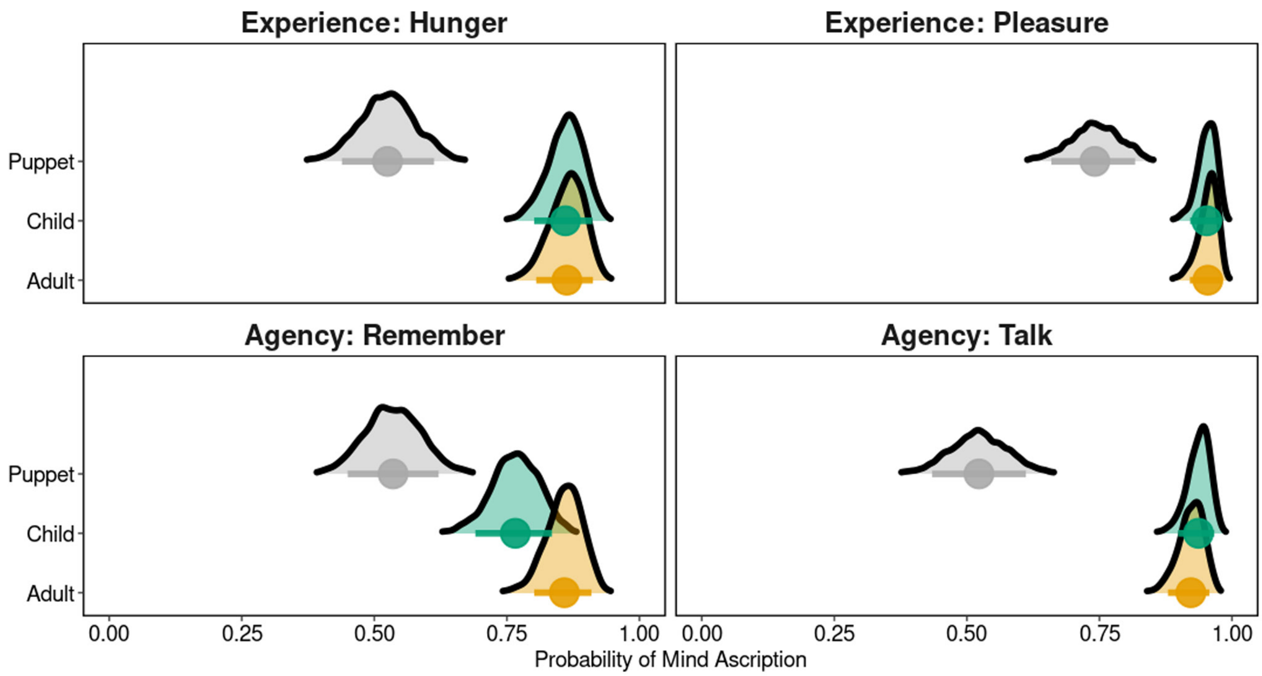

FIGURE 2 Estimated probability that a child will ascribe mind properties to puppets, adults, and peers (based on the puppet-as-puppet model to illustrate a variation across partners). Children's age and sex are set at zero. Dots present posterior means, horizontal lines 89%-highest posterior density. The upper panel shows posterior distributions for both items in the experience domain (i.e., hunger and pleasure), the lower panel for the agency domain (i.e., remember and talk).

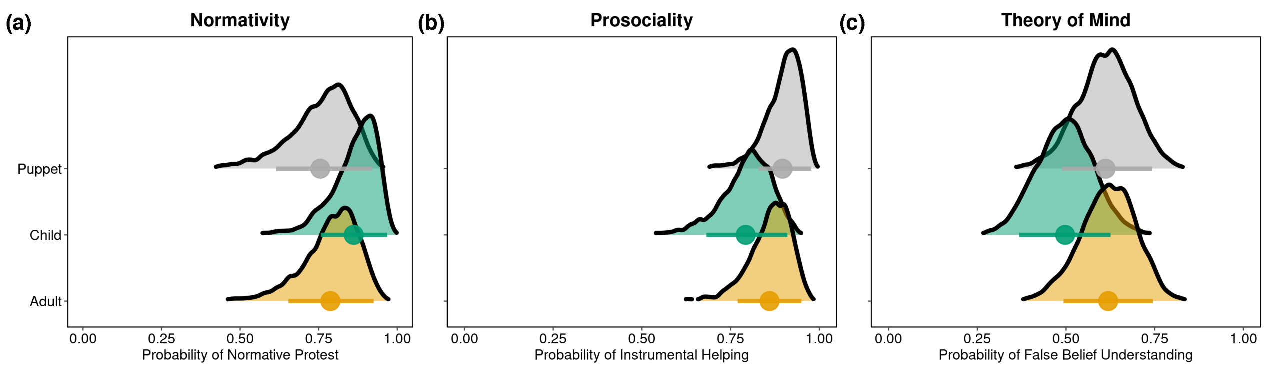

FIGURE 3 Estimated probability that a child will (a) protest (b) help, or (c) ascribe false beliefs to puppets, peers, and adult partners. Predictions obtained from puppet-as-puppet model to allow for variation across all three conditions. Children's age and sex are set at zero. Dots present posterior means, horizontal lines 89%-highest posterior density.

Fixed Effect$_{Child}$: Estimate ± SE = 0.84 ± 0.27; *Fixed Effect*$_{Adult}$: Estimate ± SE = 1.12 ± 0.28). The posterior probability that children helped puppets more like they helped peers than adults was .26.

The model's 89% HPD interval for the difference in the probability of children's instrumental helping in the puppet and CC included zero (89%-HPD = [−0.01; 0.22]). Further WAIC comparisons indicated that the parsimonious *Model*$_{no\ condition}$ (WAIC ± SE = 115.9 ± 16.6; weight = .35) had marginally better predictive accuracy than the *Model*$_{puppet\text{-}as\text{-}adult}$ (WAIC ± SE = 116.1 ± 16.2; weight = .33). Both models outperformed the *Model*$_{puppet\text{-}as\text{-}puppet}$ (WAIC ± SE = 117.4 ± 16.1; weight = .17) and the *Model*$_{puppet\text{-}as\text{-}child}$ (WAIC ± SE = 117.5 ± 16.1; weight = .16). Again, differences in WAICs across models were much smaller than the related standard errors indicating that models' predictive accuracies did not vary strongly. In line with the *Model*$_{no\ condition}$ and the *Model*$_{puppet\text{-}as\text{-}adult}$, we conclude that children helped puppets and adults alike. There remained some uncertainty as to whether children might help peers at slightly lower rates than both adults and puppets, but such effects, if present, would be rather negligible. Considering the predictive accuracy and the parsimony of the *Model*$_{no\ condition}$, we conclude that children help puppets, adults, and peers alike (see Figure 3b).

Theory of mind task

Children ascribed false beliefs to puppets (prob. = .55), peers (prob. = .55), and adults (prob. = .61) at intermediate rates. The was no indication for an effect of condition on children's false belief understanding (*Fixed Effect*$_{Puppet}$: Estimate ± SE = 0.29 ± 0.21; *Fixed Effect*$_{Child}$: Estimate ± SE = −0.01 ± 0.21; *Fixed Effect*$_{Adult}$: Estimate ± SE = 0.31 ± 0.21). The posterior probability of children ascribing false beliefs to puppets as to children compared with adults was .33.

Further analyses confirmed little variation in children's false belief understanding across conditions (see Figure 3c). The model's 89% HPD interval for the probability differences between the puppet and CCs included zero (89%-HPD = [−0.05; 0.28]). WAIC-comparisons suggested that the *Model*$_{no\ condition}$ (WAIC ± SE = 178.6 ± 10.7; weight = .40) had better predictive accuracy than the *Model*$_{puppet\text{-}as\text{-}adult}$ (WAIC ± SE = 179.1 ± 11.0; weight = .32). Again, both models outperformed the *Model*$_{puppet\text{-}as\text{-}puppet}$ (WAIC ± SE = 181.2 ± 11.2; weight = .11) and the *Model*$_{puppet\text{-}as\text{-}child}$ (WAIC ± SE = 180.3 ± 10.8; weight = .17). Once again, WAICs varied little in relation to the models' WAIC's standard errors, indicating only minor variation in predictive accuracy across models. We thus conclude that children ascribe false beliefs similarly to puppets, adults, and peers.

DISCUSSION

Given apparent methodological challenges of studying social cognition in peer interactions, scholars commonly rely on puppets, animated by adult experimenters, as stand-ins for peers (Rakoczy, 2022; Stengelin et al., 2022). However, it has been heavily debated whether children engage with puppets as if they were real-world human partners (Lillard, 2022; Packer & Moreno-Dulcey, 2022; Rakoczy, 2022). This Registered Report provides an empirical test of whether children's social cognition with puppets matches their social cognition with peers, adults, or neither. We focused on children's protest against conventional norm violations (i.e., normativity), their instrumental helping (i.e., prosociality), and their false belief understanding (i.e., theory of mind). We also assessed children's mind ascriptions (i.e., agency and experience) to puppets, children, and adults.

Our study has two key findings. First, children attributed more agency and experience to peers and adults than to puppets. Thus, we found clear support for the puppet-as-puppet hypothesis according to which children distinguish between humans and puppets in their mind ascriptions. By contrast, we found no support for the puppet-as-puppet hypothesis for children's normativity, prosociality, and theory of mind. We also found no support for the puppet-as-child hypothesis in these domains of social cognition. Models assuming that children treat all interaction partners the same or that they treat puppet and adults alike predicted the data best, indicating that children's social cognition with puppets closely resembles their social cognition with adults. Our analyses cannot fully rule out that there are slight differences between child partners compared with puppets or adults in the context of normative protest (i.e., higher protest rates with peers) and instrumental helping (i.e., lower helping rates with peers). It is important to note that, if such peer effects were true, their effect sizes were markedly small and likely beyond the minimal effect size of interest we had preregistered. Combined, these results resonate with Lillard's assumption that "children (at least by age 3) take an intentional stance towards puppets, understanding that they are not actually sentient and animate, but willing to project their theory of mind as if they were" (2022, p. 4). We not only provide direct empirical support for this claim in the context of children's false belief understanding, but also present data to extend this claim to the domains of normativity and prosociality.

Our results speak to both sides of the ongoing puppetry debate in developmental psychology: We validated puppetry as a standard practice in developmental psychology research (Kominsky et al., 2022; Rakoczy, 2022; Yu & Wellman, 2022) by showing that puppets evoke realistic social cognition in young children. Hence, we confirm the use of puppets as stand-ins for human partners (adult or peer) in developmental research across a range of socio-cognitive phenomena (normativity, prosociality, theory of mind). However, children ascribed agency and experiences less frequently to puppets than to peers and adults. This finding aligns, to some extent, with critics of puppetry who maintain that puppet studies rely on experimental tasks being framed as pretense and hence may lack ecological validity (Packer & Moreno-Dulcey, 2022). Our results suggest that children understand that puppets' agency and experiences are not human-like but interact with them as if they were humans, likely by tapping into their ability for pretend play (Lillard, 2022).

We did not find support for the puppet-as-child hypothesis in relation to the competing hypotheses for any of the domains of social cognition under investigation. This finding is noteworthy given that introducing puppets as stand-ins for peers is a common application of puppetry in developmental research. From a methodological perspective, our results suggest that the gains in ecological validity of this application of puppetry are negligible. In fact, children's social cognition with puppets resembled their social cognition with adults more than peers, but children's normative protest, instrumental helping, and false belief understanding generally varied little across adult and child partners. This raises some doubt about the necessity to use puppets as stand-ins for peers when investigating these phenomena.

At first glance, this pattern of results opposes previous work documenting differences in social cognition when interacting with peers compared with adults (Seehagen et al., 2018; Ulber & Tomasello, 2020). However, our study protocol and research questions differed somewhat from the previous. For example, Ulber and Tomasello (2020) emphasized the contrast between the adult and peer partner by introducing adults formally with their last name. Children may interact differently with adults and peers if social hierarchies are accentuated, but these differences likely shrink once adult partners are introduced as equals. To simulate children's social cognition in peer interactions in developmental science, researchers may equally well introduce adults instead of puppets.

According to our findings, puppetry provides valid insights into childhood social cognition under certain circumstances. Before generalizing these results to childhood social cognition research more broadly, however, it is important to consider the procedural details of our study: We investigated specific phenomena of social cognition (i.e., normative protest, instrumental helping, and false belief understanding) among 3- and 4-year-old German preschoolers in a face-to-face setting. Thus, constraints on generality (Simons et al., 2017) may apply depending on (i) participants' cultural background, (ii) their ages, (iii) the socio-cognitive phenomena under investigation, and (iv) the social setting in which puppets are employed.

First, it is important to note that the current study tested puppetry in a particular socio-cultural setting: German children grow up with an abundance of puppets, dolls, mass-manufactured artifacts, and cartoon characters. Cultural variation in these types of experiences may also affect how readily children (temporarily) accept them as animate agents within pretend play. Indeed, children's pretense tendencies vary across cultures (Gaskins, 2013), raising some doubt about the cross-cultural application of puppetry. On the contrary, a recent meta-analysis reported little variation across countries in children's false belief ascriptions to humans as compared to puppets, dolls, or other artifacts (Yu & Wellman, 2022; see also Wellman et al., 2001). Their finding thus indicates some cultural robustness in the use of puppetry in theory of mind research. Furthermore, German children are socialized to be psychologically autonomous and to freely enter and navigate social interactions with adult and peer partners (Keller & Kärtner, 2013), while other societies emphasize social hierarchies between children and adult caregivers (Keller, 2016). In these cultural contexts, children may be more sensitive to their partner's age and social position and, as a result, exhibit more variation in their social cognition across partners. In any case, scholars need to consider and describe the socio-cultural variables of their sampled communities when relying on and interpreting puppetry research.

Secondly, the value of puppetry in developmental research may likely vary with participants' ages. Here, we investigated early preschoolers from a rather narrow age range of three to 4 years. In a recent study, Stengelin et al. (2022) found support for the *puppet-as-puppet* hypothesis in the context of social learning among slightly older children from the same community. It is possible that children become increasingly attuned to partner characteristics during their preschool years. The current age range appears ideally suited for applying puppetry in developmental research (see also Bartsch et al., 2011; Lillard, 2022), but older children may exhibit different social cognition when interacting with puppets compared with human partners. If so, applying puppets across wide age ranges might introduce (unwanted) confounds that need to be acknowledged and accounted for.

A third constraint regarding the generalizability of the current findings concerns the socio-cognitive phenomena under investigation. For example, Stengelin et al. (2022) found support for the puppet-as-puppet hypothesis in children's overimitation, a social learning strategy that is known to fluctuate across social contexts (Hoehl et al., 2019). Interestingly, in both the current study and the study by Stengelin et al. (2022), children's behaviors with puppets more closely resembled those with adults rather than peers. However, children overimitated adults more than child models, which may have also fostered variation in overimitation of child and puppet models (Stengelin et al., 2021). Accordingly, puppetry may be well-suited to simulate peer interactions for robust phenomena of social cognition, but less adequate for studying more context-sensitive phenomena such as children's overimitation (Stengelin et al., 2022; Wood et al., 2012) or other aspects of selective social learning (e.g., Jaswal & Neely, 2006; Kachel, Moore, et al., 2018). Moreover, puppets may be less suitable to explore questions where children's mind or emotional ascriptions to social partners are central, such as their reasoning about whether someone feels guilty about violating moral, rather than conventional norms. In the domain of prosociality, one may speculate that children could be more sensitive to partners' identities when deciding whether to comfort or empathize with someone in distress. More fine-grained (or repeated) measures than the binary measures we used in our study may also reveal qualitative differences in how children engage with puppets, adults, or peers (e.g., Jaswal & Neely, 2006; Mammen et al., 2019; Stengelin et al., 2022).

Finally, the use of puppets may vary across social settings. Here, we applied puppets in a live (face-to-face) setting and found, if any, only subtle variation in social cognition across partners. Other research has tested puppetry in mock video calls in which children and their partners were spatially separated (Stengelin et al., 2022). Such paradigms allow for reciprocal communication while maintaining rigorous standardization (see also Kachel et al., 2021), but the social demands and affordances of video versus face-to-face settings may nevertheless lead to variation in children's social cognition when engaging with puppets compared with human partners (e.g., Nielsen et al., 2008). Once again, these points would need to be taken into account when generalizing the current findings to puppetry when applied in other social settings.

Given these considerations, we want to emphasize that puppetry presents only one among many potential approaches to studying social cognition in peer interactions. Alternative approaches such as peer confederates (Engelmann et al., 2018) or dyadic peer paradigms (Kachel, Moore, et al., 2018) are viable alternatives to puppetry as they rely less on children's pretense but instead observe ongoing interactions in naturalistic settings. Furthermore, such approaches have proven useful in the context of cross-cultural research on childhood social cognition (Kanngiesser et al., 2022; Schäfer et al., 2015; Zeidler et al., 2016) and may thus allow for better generalizability of findings outside the study context.

Taken together, we show that children ascribe less agency and experiences to puppets than to human partners, both adults and peers. At the same time, children's social cognition, such as their normativity, prosociality, and theory of mind, appears robust to whether they engage with puppets or human partners. We argue that developmental researchers need to better justify their assumptions when employing puppets as stand-ins for human partners. Puppets are widespread in developmental psychology, but it is rarely explicitly justified whether puppets are used to simulate peers, humans of all ages, or animate agents more broadly. Moreover, researchers employ puppets in diverse forms and appearances, including single-hand animal puppets or puppets with child-like appearances as in the current study (see also Lillard, 2022). It will be important for future research to better communicate and justify such choices to enable direct replications and adequate generalizations based on puppetry research.

ACKNOWLEDGMENTS

We would like to thank Jana Jurkat, Felix Buech, Victoria Gassner, Nane Herfurth, Moritz Schmid, Christoph Nostitz, Stefan Gregor, Ljubica Petrovic, Linda Schymanski, Luke Maurits, and Rabea Ball for their support. We would further like to thank the reviewers at both stages of the submission process. Finally, we would like to thank all participating children and their caregivers. Open Access funding enabled and organized by Projekt DEAL.

FUNDING INFORMATION

This research was funded by the Max Planck Society.

CONFLICT OF INTEREST STATEMENT

The authors have no competing interests to declare.

DATA AVAILABILITY STATEMENT

The analyses presented here were preregistered (Registered Report). Data and code necessary to reproduce the analyses are publicly accessible, as are the materials needed to replicate the findings. All information is available at the Open Science Framework (https://osf.io/wv8mq/).

ORCID

Roman Stengelin https://orcid.org/0000-0003-2212-4613
Patricia Kanngiesser https://orcid.org/0000-0003-1068-3725

REFERENCES

Anderson, D. R., & Pempek, T. A. (2005). Television and very young children. *American Behavioral Scientist*, *48*(5), 505–522. https://doi.org/10.1177/0002764204271506

Asaba, M., Li, X., Yow, W. Q., & Gweon, H. (2022). Children selectively demonstrate their competence to a puppet when others depict it as an agent. *Cognitive Development*, *62*, 101186. https://doi.org/10.1016/j.cogdev.2022.101186

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, *21*, 37–46. https://doi.org/10.1016/0010-0277(85)90022-8

Bartsch, K., Wade, C. E., & Estes, D. (2011). Children's attention to others' beliefs during persuasion: Improvised and selected arguments to puppets and people: Attention to beliefs during persuasion. *Social Development*, *20*(2), 316–333. https://doi.org/10.1111/j.1467-9507.2010.00580.x

Benenson, J. F. (1993). Greater preference among females than males for dyadic interaction in early childhood. *Child Development*, *64*(2), 544–555. https://doi.org/10.1111/j.1467-8624.1993.tb02927.x

Bohn, M., Kordt, C., Braun, M., Call, J., & Tomasello, M. (2020). Learning novel skills from iconic gestures: A developmental and evolutionary perspective. *Psychological Science*, *31*, 873–880. https://doi.org/10.1177/0956797620921519

Broesch, T., Carolan, P. L., Cebioğlu, S., von Rueden, C., Boyette, A., Moya, C., Hewlett, B., & Kline, M. A. (2021). Opportunities for interaction: Natural observations of children's social behavior in five societies. *Human Nature*, *32*, 208–238. https://doi.org/10.1007/s12110-021-09393-w

Brownell, C. A., Ramani, G. B., & Zerwas, S. (2006). Becoming a social partner with peers: Cooperation and social understanding in one- and two-year-olds. *Child Development*, *77*(4), 803–821. https://doi.org/10.1111/j.1467-8624.2006.00904.x

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, *10*(1), 395. https://doi.org/10.32614/RJ-2018-017

Callaghan, T., Moll, H., Rakoczy, H., Warneken, F., Liszkowski, U., Behne, T., & Tomasello, M. (2011). *Early social cognition in three cultural contexts* (Vol. *76*, 2nd ed.). Wiley-Blackwell.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, *13*(4), 148–153. https://doi.org/10.1016/j.tics.2009.01.005

Damiano, L., & Dumouchel, P. (2018). Anthropomorphism in human-robot co-evolution. *Frontiers in Psychology*, *9*, 468. https://doi.org/10.3389/fpsyg.2018.00468

Dolgin, K. G., & Behrend, D. A. (1984). Children's knowledge about animates and Inanimates. *Child Development*, *55*(4), 1646. https://doi.org/10.2307/1130034

Dorie, B., Tranby, Z., Van Cleave, S., Cardella, M., & Svarovsky, G. (2013). *Using puppets to elicit talk during interviews on engineering with young children*. 2013 ASEE Annual Conference & Exposition Proceedings, 23.1334.1–23.1334.7. https://doi.org/10.18260/1-2--22719

Dunfield, K. A., Kuhlmeier, V. A., & Murphy, L. (2013). Children's use of communicative intent in the selection of cooperative partners. *PLoS One*, *8*(4), e61804. https://doi.org/10.1371/journal.pone.0061804

Engelmann, J. M., Herrmann, E., Rapp, D. J., & Tomasello, M. (2016). Young children (sometimes) do the right thing even when their peers do not. *Cognitive Development*, *39*, 86–92. https://doi.org/10.1016/j.cogdev.2016.04.004
Engelmann, J. M., Herrmann, E., & Tomasello, M. (2018). Concern for group reputation increases prosociality in young children. *Psychological Science*, *29*(2), 181–190. https://doi.org/10.1177/0956797617733830
Epstein, I., Stevens, B., McKeever, P., Baruchel, S., & Jones, H. (2008). Using puppetry to elicit children's talk for research. *Nursing Inquiry*, *15*(1), 49–56. https://doi.org/10.1111/j.1440-1800.2008.00395.x
Gampe, A., & Daum, M. M. (2018). How preschoolers react to norm violations is associated with culture. *Journal of Experimental Child Psychology*, *165*, 135–147. https://doi.org/10.1016/j.jecp.2017.06.009
Gaskins, S. (2013). *Pretend play as culturally constructed activity*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195395761.013.0016
Gauvain, M., & Rogoff, B. (1989). Collaborative problem solving and children's planning skills. *Developmental Psychology*, *25*(1), 139–151. https://doi.org/10.1037/0012-1649.25.1.139
Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development*, *59*(1), 26. https://doi.org/10.2307/1130386
Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*(5812), 619. https://doi.org/10.1126/science.1134475
Green, M., Kirby, J. N., & Nielsen, M. (2018). The cost of helping: An exploration of compassionate responding in children. *British Journal of Developmental Psychology*, *36*(4), 673–678. https://doi.org/10.1111/bjdp.12252
Grueneisen, S., Duguid, S., Saur, H., & Tomasello, M. (2017). Children, chimpanzees, and bonobos adjust the visibility of their actions for cooperators and competitors. *Scientific Reports*, *7*(1), 8504. https://doi.org/10.1038/s41598-017-08435-7
Hardecker, S., Schmidt, M. F. H., Roden, M., & Tomasello, M. (2016). Young children's behavioral and emotional responses to different social norm violations. *Journal of Experimental Child Psychology*, *150*, 364–379. https://doi.org/10.1016/j.jecp.2016.06.012
Hepach, R., Kante, N., & Tomasello, M. (2017). Toddlers help a peer. *Child Development*, *88*(5), 1642–1652. https://doi.org/10.1111/cdev.12686
Hewlett, B. S. (1996). *Diverse contexts of human infancy*. Prentice Hall.
Heyman, G. D., Chiu Loke, I., & Lee, K. (2016). Children spontaneously police adults' transgressions. *Journal of Experimental Child Psychology*, *150*, 155–164. https://doi.org/10.1016/j.jecp.2016.05.012
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. https://doi.org/10.1016/j.dr.2018.12.002
Huber, L., Plötner, M., & Schmitz, J. (2019). Behavioral observation of prosocial behavior and social initiative is related to preschoolers' psychopathological symptoms. *PLoS One*, *14*(11), e0225274. https://doi.org/10.1371/journal.pone.0225274
Jaswal, V. K., & Neely, L. A. (2006). Adults don't always know best: Preschoolers use past reliability over age when learning new words. *Psychological Science*, *17*(9), 757–758. https://doi.org/10.1111/j.1467-9280.2006.01778.x
Jipson, J. L., & Gelman, S. A. (2007). Robots and rodents: Children's inferences about living and nonliving kinds. *Child Development*, *78*(6), 1675–1688. https://doi.org/10.1111/j.1467-8624.2007.01095.x
Kachel, G., Hepach, R., Moore, R., & Tomasello, M. (2021). Two- but not three-year-olds prefer adults as informants in an object-choice-task. *Child Development*, *92*, e635–e652. https://doi.org/10.1111/cdev.13544
Kachel, G., Moore, R., & Tomasello, M. (2018). Two-year-olds use adults' but not peers' points. *Developmental Science*, *21*, e12660. https://doi.org/10.1111/desc.12660
Kachel, U., Svetlova, M., & Tomasello, M. (2018). Three-year-olds' reactions to a partner's failure to perform her role in a joint commitment. *Child Development*, *89*(5), 1691–1703. https://doi.org/10.1111/cdev.12816
Kanngiesser, P., Schäfer, M., Herrmann, E., Zeidler, H., Haun, D., & Tomasello, M. (2022). Children across societies enforce conventional norms but in culturally variable ways. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(1), e2112521118. https://doi.org/10.1073/pnas.2112521118
Kanngiesser, P., & Warneken, F. (2012). Young children consider merit when sharing resources with others. *PLoS One*, *7*(8), e43979. https://doi.org/10.1371/journal.pone.0043979
Keller, H. (2007). *Cultures of infancy*. Lawrence Erlbaum Associates.
Keller, H. (2016). Psychological autonomy and hierarchical relatedness as organizers of developmental pathways. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1686), 20150070. https://doi.org/10.1098/rstb.2015.0070
Keller, H., & Kärtner, J. (2013). The cultural solution of universal developmental tasks. *Advances in Culture and Psychology*, *3*, 63–116.
Kim, S., Sodian, B., & Paulus, M. (2014). "Does he need help or can he help himself?" preschool children's expectations about others' instrumental helping versus self-helping. *Frontiers in Psychology*, *5*, 430. https://doi.org/10.3389/fpsyg.2014.00430
Kline, M. A. (2015). How to learn about teaching: An evolutionary framework for the study of teaching behavior in humans and other animals. *Behavioral and Brain Sciences*, *38*, e31. https://doi.org/10.1017/S0140525X14000090
Kominsky, J. F., Lucca, K., Thomas, A. J., Frank, M. C., & Hamlin, J. K. (2022). Simplicity and validity in infant research. *Cognitive Development*, *63*, 101213. https://doi.org/10.1016/j.cogdev.2022.101213
Köymen, B., Lieven, E., Engemann, D. A., Rakoczy, H., Warneken, F., & Tomasello, M. (2014). Children's norm enforcement in their interactions with peers. *Child Development*, *85*(3), 1108–1122. https://doi.org/10.1111/cdev.12178
Lancy, D. F. (2008). *The anthropology of childhood: Cherubs, chattel, changelings*. Cambridge University Press.
Lewis, M., Feiring, C., & Kotsonis, M. (1984). The social network of the young child. In M. Lewis (Ed.), *Beyond the dyad* (pp. 129–160). Springer US. https://doi.org/10.1007/978-1-4757-9415-1_7
Lew-Levy, S., Kissler, S. M., Boyette, A. H., Crittenden, A. N., Mabulla, I. A., & Hewlett, B. S. (2019). Who teaches children to forage? Exploring the primacy of child-to-child teaching among Hadza and BaYaka hunter-gatherers of Tanzania and Congo. *Evolution and Human Behavior*, *41*, 12–22. https://doi.org/10.1016/j.evolhumbehav.2019.07.003
Lillard, A. S. (2022). Pretending at hand: How children perceive and process puppets. *Cognitive Development*, *63*, 101202. https://doi.org/10.1016/j.cogdev.2022.101202
Mammen, M., Köymen, B., & Tomasello, M. (2019). Children's reasoning with peers and parents about moral dilemmas. *Developmental Psychology*, *55*(11), 2324–2335. https://doi.org/10.1037/dev0000807
McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan*. CRC Press.
Moll, H., Khalulyan, A., & Moffett, L. (2017). 2.5-year-olds express suspense when others approach reality with false expectations. *Child Development*, *88*(1), 114–122. https://doi.org/10.1111/cdev.12581
Molleman, L., Kanngiesser, P., & van den Bos, W. (2019). Social information use in adolescents: The impact of adults, peers and household composition. *PLoS One*, *14*(11), e0225498. https://doi.org/10.1371/journal.pone.0225498

Moriguchi, Y., Takahashi, H., Nakamata, T., & Todo, N. (2019). Mind perception of god in Japanese children. *International Journal of Psychology*, *54*(4), 557–562. https://doi.org/10.1002/ijop.12482

Nielsen, M., Haun, D., Kärtner, J., & Legare, C. H. (2017). The persistent sampling bias in developmental psychology: A call to action. *Journal of Experimental Child Psychology*, *162*, 31–38. https://doi.org/10.1016/j.jecp.2017.04.017

Nielsen, M., & Haun, D. B. M. (2016). Why developmental psychology is incomplete without comparative and cross-cultural perspectives. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1686), 20150071. https://doi.org/10.1098/rstb.2015.0071

Nielsen, M., Simcock, G., & Jenkins, L. (2008). The effect of social engagement on 24-month-olds' imitation from live and televised models. *Developmental Science*, *11*(5), 722–731. https://doi.org/10.1111/j.1467-7687.2008.00722.x

Packer, M. J., & Moreno-Dulcey, F. A. (2022). Theory of puppets?: A critique of the use of puppets as stimulus materials in psychological research with young children. *Cognitive Development*, *61*, 101146. https://doi.org/10.1016/j.cogdev.2021.101146

Plötner, M., Over, H., Carpenter, M., & Tomasello, M. (2015). The effects of collaboration and minimal-group membership on children's prosocial behavior, liking, affiliation, and trust. *Journal of Experimental Child Psychology*, *139*, 161–173. https://doi.org/10.1016/j.jecp.2015.05.008

R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Radziszewska, B., & Rogoff, B. (1988). Influence of adult and peer collaborators on children's planning skills. *Developmental Psychology*, *24*(6), 840–848. https://doi.org/10.1037/0012-1649.24.6.840

Radziszewska, B., & Rogoff, B. (1991). Children's guided participation in planning imaginary errands with skilled adult or peer partners. *Developmental Psychology*, *27*(3), 381–389. https://doi.org/10.1037/0012-1649.27.3.381

Rakoczy, H. (2022). Puppet studies present clear and distinct windows into the child's mind. *Cognitive Development*, *61*, 101147. https://doi.org/10.1016/j.cogdev.2021.101147

Rakoczy, H., Brosche, N., Warneken, F., & Tomasello, M. (2009). Young children's understanding of the context-relativity of normative rules in conventional games. *British Journal of Developmental Psychology*, *27*(2), 445–456. https://doi.org/10.1348/026151008X337752

Rakoczy, H., Hamann, K., Warneken, F., & Tomasello, M. (2010). Bigger knows better: Young children selectively learn rule games from adults rather than from peers. *British Journal of Developmental Psychology*, *28*(4), 785–798. https://doi.org/10.1348/026151009X479178

Rakoczy, H., & Schmidt, M. F. H. (2013). The early ontogeny of social norms. *Child Development Perspectives*, *7*(1), 17–21. https://doi.org/10.1111/cdep.12010

Rakoczy, H., Warneken, F., & Tomasello, M. (2008). The sources of normativity: Young children's awareness of the normative structure of games. *Developmental Psychology*, *44*(3), 875–881. https://doi.org/10.1037/0012-1649.44.3.875

Reiß, M., Krüger, M., & Krist, H. (2019). Theory of mind and the video deficit effect: Video presentation impairs Children's encoding and understanding of false belief. *Media Psychology*, *22*(1), 23–38. https://doi.org/10.1080/15213269.2017.1412321

Rekers, Y., Haun, D. B. M., & Tomasello, M. (2011). Children, but not chimpanzees, prefer to collaborate. *Current Biology*, *21*(20), 1756–1758. https://doi.org/10.1016/j.cub.2011.08.066

Remer, R., & Tzuriel, D. (2015). "I teach better with the puppet"—Use of puppet as a mediating tool in kindergarten education—An evaluation. *American Journal of Educational Research*, *3*(3), 356–365. https://doi.org/10.12691/education-3-3-15

Revencu, B., & Csibra, G. (2020). Opening the black box of early depiction interpretation: From whether to how in the theory-of-puppets debate. Commentary on Kominsky, Lucca, Thomas, Frank, & Hamlin (submitted) [preprint]. *PsyArXiv*. https://doi.org/10.31234/osf.io/u3em8

Rogoff, B. (2003). *The cultural nature of human development*. Oxford University Press.

Rogoff, B. (2016). Culture and participation: A paradigm shift. *Current Opinion in Psychology*, *8*, 182–189. https://doi.org/10.1016/j.copsyc.2015.12.002

Rubio-Fernández, P., & Geurts, B. (2013). How to pass the false-belief task before your fourth birthday. *Psychological Science*, *24*(1), 27–33. https://doi.org/10.1177/0956797612447819

Rudolph, K. D., & Heller, T. L. (1997). Interpersonal problem solving, externalizing behavior, and social competence in preschoolers: A knowledge-performance discrepancy? *Journal of Applied Developmental Psychology*, *18*(1), 107–117. https://doi.org/10.1016/S0193-3973(97)90017-4

Schäfer, M., Haun, D. B., & Tomasello, M. (2015). Fair is not fair everywhere. *Psychological Science*, *26*(8), 1252–1260. https://doi.org/10.1177/0956797615586188

Schmidt, M. F. H., Hardecker, S., & Tomasello, M. (2016). Preschoolers understand the normativity of cooperatively structured competition. *Journal of Experimental Child Psychology*, *143*, 34–47. https://doi.org/10.1016/j.jecp.2015.10.014

Seehagen, S., Dreier, L., & Zmyj, N. (2018). Overrated adults: 4-year-olds' false belief understanding is influenced by the believer's age. *Journal of Experimental Child Psychology*, *167*, 328–335. https://doi.org/10.1016/j.jecp.2017.11.007

Simon, S., Naylor, S., Keogh, B., Maloney, J., & Downing, B. (2008). Puppets promoting engagement and talk in science. *International Journal of Science Education*, *30*(9), 1229–1248. https://doi.org/10.1080/09500690701474037

Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on generality (COG): A proposed addition to all empirical papers. *Perspectives on Psychological Science*, *12*(6), 1123–1128. https://doi.org/10.1177/1745691617708630

Stengelin, R., Ball, R., Maurits, L., Kanngiesser, P., & Haun, D. B. M. (2022). Children overimitate adults and peers more than puppets. *Developmental Science*, e13303. https://doi.org/10.1111/desc.13303

Stengelin, R., Grueneisen, S., & Tomasello, M. (2018). Why should I trust you? Investigating young children's spontaneous mistrust in potential deceivers. *Cognitive Development*, *48*, 146–154. https://doi.org/10.1016/j.cogdev.2018.08.006

Stengelin, R., Haun, D. B. M., & Kanngiesser, P. (2021). *Simulating peers: Can puppets simulate peer interactions in studies on children's socio-cognitive development?* 10.17605/OSF.IO/DQYCM

Stengelin, R., Hepach, R., & Haun, D. B. M. (2020). Cultural variation in young children's social motivation for peer collaboration and its relation to the ontogeny of theory of mind. *PLoS One*, *15*(11), e0242071. https://doi.org/10.1371/journal.pone.0242071

Sutherland, S. L., & Friedman, O. (2012). Preschoolers acquire general knowledge by sharing in pretense: Learning from pretense. *Child Development*, *83*(3), 1064–1071. https://doi.org/10.1111/j.1467-8624.2012.01748.x

Taylor, M., Cartwright, B. S., & Bowden, T. (1991). Perspective taking and theory of mind: Do children predict interpretive diversity as a function of differences in Observers' knowledge? *Child Development*, *62*(6), 1334–1351. https://doi.org/10.2307/1130810

Terashima, H., & Hewlett, B. S. (Eds.). (2016). *Social learning and innovation in contemporary hunter-gatherers*. Springer Japan. https://doi.org/10.1007/978-4-431-55997-9

Thompson, B. N., & Goldstein, T. R. (2020). Children learn from both embodied and passive pretense: A replication and extension. *Child Development*, *91*(4), 1364–1374.

Tomasello, M. (2019). *Becoming human: A theory of ontogeny*. Harvard University Press.

Tomasello, M., & Gonzalez-Cabrera, I. (2017). The role of ontogeny in the evolution of human cooperation. *Human Nature*, *28*(3), 274–288. https://doi.org/10.1007/s12110-017-9291-1

Ulber, J., & Tomasello, M. (2020). Young children's prosocial responses towards peers and adults in two social contexts. *Journal of Experimental Child Psychology*, *198*, 104888. https://doi.org/10.1016/j.jecp.2020.104888

Urquiza-Haas, E. G., & Kotrschal, K. (2015). The mind behind anthropomorphic thinking: Attribution of mental states to other species. *Animal Behaviour*, *109*, 167–176. https://doi.org/10.1016/j.anbehav.2015.08.011

Vaish, A., Missana, M., & Tomasello, M. (2011). Three-year-old children intervene in third-party moral transgressions: Children intervene in moral transgressions. *British Journal of Developmental Psychology*, *29*(1), 124–130. https://doi.org/10.1348/026151010X532888

van Leeuwen, E. J. C., Cohen, E., Collier-Baker, E., Rapold, C. J., Schäfer, M., Schütte, S., & Haun, D. B. M. (2018). The development of human social learning across seven societies. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-04468-2

VanderBorght, M., & Jaswal, V. K. (2009). Who knows best? Preschoolers sometimes prefer child informants over adult informants. *Infant and Child Development*, *18*(1), 61–71. https://doi.org/10.1002/icd.591

Walker-Andrews, A. S., & Kahana-Kalman, R. (1999). The understanding of pretence across the second year of life. *British Journal of Developmental Psychology*, *17*(4), 523–536. https://doi.org/10.1348/026151099165456

Warneken, F., & Tomasello, M. (2013). The emergence of contingent reciprocity in young children. *Journal of Experimental Child Psychology*, *116*(2), 338–350. https://doi.org/10.1016/j.jecp.2013.06.002

Weisman, K., Dweck, C. S., & Markman, E. M. (2018). Folk philosophy of mind: Changes in conceptual structure between 4–9 y of age. *Proceedings of the 40th Annual Conference of the Cognitive Science Society*, 6.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, *72*(3), 655–684.

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

Wood, L. A., Kendal, R. L., & Flynn, E. G. (2012). Context-dependent model-based biases in cultural transmission: Children's imitation is affected by model age over model knowledge state. *Evolution and Human Behavior*, *33*(4), 387–394. https://doi.org/10.1016/j.evolhumbehav.2011.11.010

Yu, C.-L., & Wellman, H. M. (2022). Young children treat puppets and dolls like real persons in theory of mind research: A meta-analysis of false-belief understanding across ages and countries. *Cognitive Development*, *63*, 101197. https://doi.org/10.1016/j.cogdev.2022.101197

Zeidler, H., Herrmann, E., Haun, D. B. M., & Tomasello, M. (2016). Taking turns or not? Children's approach to limited resource problems in three different cultures. *Child Development*, *87*(3), 677–688. https://doi.org/10.1111/cdev.12505

Zmyj, N., Daum, M. M., Prinz, W., Nielsen, M., & Aschersleben, G. (2012). Fourteen-month-olds' imitation of differently aged models: Imitation of differently aged models. *Infant and Child Development*, *21*(3), 250–266. https://doi.org/10.1002/icd.750

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Stengelin, R., Haun, D. B. M., & Kanngiesser, P. (2023). Simulating peers: Can puppets simulate peer interactions in studies on children's socio-cognitive development? *Child Development*, *00*, 1–19. https://doi.org/10.1111/cdev.13913