

# PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

BIOLOGICAL SCIENCES

## How two- and four-year-old children coordinate joint actions with peers

Journal:	<i>Philosophical Transactions B</i>
Manuscript ID	RSTB-2021-0100.R1
Article Type:	Research
Date Submitted by the Author:	13-Apr-2022
Complete List of Authors:	Rossano, Federico; University of California San Diego, Department of Cognitive Science Terwilliger, Jack; University of California San Diego, Department of Cognitive Science Bangerter, A; University of Neuchâtel, Institute of Work and Organizational Psychology Genty, Emilie; Universite de Neuchatel, Biology institute Heesen, Raphaela; Durham University, Department of Psychology ZUBERBÜHLER, Klaus; Universite de Neuchatel, Institute of biology
Issue Code (this should have already been entered and appear below the blue box, but please contact the Editorial Office if it is not present):	INTERACTION
Subject:	Behaviour < BIOLOGY, Cognition < BIOLOGY
Keywords:	Joint Action, Children, Coordination, Interaction Engine, Social Relationships

SCHOLARONE™  
Manuscripts

## How two- and four-year-old children coordinate joint actions with peers

Federico Rossano<sup>1</sup>, Jack Terwilliger<sup>1</sup>, Adrian Bangerter<sup>2</sup>, Emilie Genty<sup>2,3</sup>, Raphaela Heesen<sup>4</sup>, Klaus Zuberbühler<sup>3,5</sup>

1. Department of Cognitive Science, University of California, San Diego, USA

2. Institute of Work and Organizational Psychology, University of Neuchâtel, Switzerland

3. Institute of Biology, University of Neuchâtel, Switzerland

4. Department of Psychology, Durham University, UK

5. School of Psychology and Neuroscience, University of St Andrews, UK

### Abstract

The interaction engine hypothesis postulates that humans have a unique ability and motivation for social interaction. A crucial juncture in the ontogeny of the interaction engine could be around 2-4 years of age, but observational studies of children in natural contexts are limited. These data appear critical also for comparison with non-human primates. Here we report on focal observations on 31 children aged two and four years old in four preschools (10h per child). Children interact with a wide range of partners, many infrequently, but with one or two close friends. Four year olds engage in cooperative joint actions more often than two year olds, and fight less than two year olds. Conversations and playing with objects are the most frequent joint action types in both age groups. Children engage in joint actions with peers frequently (on average 13 distinct joint actions per hour) and briefly (28s on average) and shorter than those of great apes in comparable studies. Their joint actions feature entry and exit phases about two-thirds of the time, less frequently than great apes. The results support the interaction engine hypothesis, as young children manifest a remarkable motivation and ability for fast-paced interactions with multiple partners.

## 1. Introduction

It has long been argued that social interaction is “the primordial site of human sociality” [1] and that contrary to popular belief, one of the most remarkable capacity of human beings is not language, but rather a special propensity to interact with one another, with no equivalent in other animal species, may it be in quantity or quality [2, 3]. The claim goes as far as suggesting that language, our ability to cooperate with each other and cultural accumulation are all crucially dependent on what has been called an “interaction engine” [2, 3]. Notably, this “engine” is not a module in the brain or a single cognitive mechanism, but rather a set of proclivities with different phylogenetic origins and ontogenetic developments that drives the way humans interact with each other and that appear to be fundamentally universal, i.e., a property of the species more than the byproduct of socialization.

Joint activities are the natural home in which the interaction engine displays its strengths [4, 5]. When two or more individuals come to coordinate their actions towards achieving a shared goal (that no individual could achieve alone), each individual is faced with the challenge of considering not just how their own action will help achieve their goal, but also the need to monitor the actions of their social partners, make predictions about them, coordinate and possibly correct or adjust their own conduct. At a very basic level, joint actions require joint attention (knowing what others can and cannot perceive); predicting what others will do, based on their ongoing behaviour but also based on past interactions, knowing what others should do to achieve the common goal [6]. According to some, this requires a matching of plans and subplans [7, 5] that on the surface appear to pose major cognitive demands and that some would claim is observable in children only around five years of age e.g., [8]. Yet, although engaging in joint actions might present complex cognitive challenges, the average human engages in joint actions for several hours a day (conversation is a type of joint action), and children as young as two year olds appear to be able to coordinate their actions with others (see e.g., [9]). It has thus been argued that joint action coordination might not necessarily rely on complex cognitive abilities but could be achieved through alignment of behaviours, allowing for a comparison to non-human animals social interactions towards investigating its evolutionary origins.

In this paper we would like to propose a move analogous to the distinction made by [10] between joint commitment as product and as process. Where joint commitment-as-product refers to the feeling of normative obligation to each other and to completing the joint action that participants might experience, while joint commitment-as-process refers to the exchange of signals between participants. In other words, the sense of commitment (the product) emerges from the signal exchange (the process) [11]. To this end, rather than focusing on the cognitive computations necessary to achieve joint actions, we want to focus on the behavioral exchange of signals between young children to document when and how they appear to behave in a more adult like manner and how they compare to our closest living relatives (chimpanzees and bonobos). From a behavioral point of view, engaging in joint actions usually entails the sequential unfolding of three phases with recognizable communicative signals: 1 an entry phase, 2 a main body and 3 an exit phase [4]. Entry phases represent the process through which a child recruits another towards engaging in a joint action while exit phases are those in which a child takes leave from another by signaling an intention to leave. The occurrence of these phases is something that has been documented in other species. For example, in a recent paper, [12] documented the presence and duration of entry and exit phases in play and grooming interactions in chimpanzees and bonobos. While entry phases occurred in 90% of bonobo interactions, they occurred in 69% of chimpanzee interactions. Entry phases in both species, lasted around 12

1  
2  
3 seconds on average. Exit phases occurred in 92% of bonobo interactions, compared with 86% of  
4 chimpanzee interactions and lasted on average around 14-17 seconds. Moreover, bonobos with  
5 closer relationships were less likely to produce entry and exit signals than those with more  
6 distant relationships, in line with what has been claimed for adult humans [13].

7  
8 How do young children manage such coordination and how do they compare to our  
9 closest living relatives? Does their ability to engage in joint actions change over time? And how  
10 do their relationships affect the coordination process?

11 In recent years there have been multiple experimental investigations of joint activities in  
12 young children, focusing on the capacity to engage in joint attention (see e.g., [14]), in social  
13 games [15] their perception of joint actions as joint commitments [16, 17, 18, 19, 20] and  
14 recognizing joint intentions [21]. While the majority of these studies entailed having a child  
15 interacting with an adult experimenter, studies looking at interactions between peers show that it  
16 is only around age two that children display an ability to coordinate with others (see e.g., [22,  
17 23]. The general findings from the experimental literature therefore suggest that between age two  
18 and three children undergo some major change in their appreciation of the social nature of joint  
19 activities and the norms and obligations that come with it. In particular, it is around that age that  
20 children orient to the existence of a joint commitment once they are engaged in an interaction  
21 with a partner. This appreciation of joint commitment can be observed behaviorally in their  
22 attempts to re-engage their partners if the joint action is suddenly interrupted, in their willingness  
23 to continue the task until both participants have obtained their reward and, in their tendency to  
24 account for their departure from the current joint task, to move on to another task with another  
25 social partner. Interestingly, a normative turn between age two and three is true not just for joint  
26 activities but several other domains such as their understanding of rules of games [24, 25],  
27 property concerns (e.g., [26]) and the occurrence of overimitation (e.g., [27]). This would  
28 suggest a maturational change occurring around that age, inviting further investigations in  
29 naturalistic settings of how young children engage in joint actions, specifically before and after 3  
30 years of age.  
31  
32  
33

34 While most recent developmental research on joint actions in children has focused on  
35 experimental paradigms, earlier observational studies have focused on young children interacting  
36 with peers in natural settings (see e.g., [28, 29]). One key constraint of most of that research was  
37 the limited ability to rely on video footage for data collection and the reliance on ethograms and  
38 observers taking notes live. As such the vast majority of studies report on data per child that  
39 amounts to maximum 1 hour, often no more than 5-10 minutes of observations collected at  
40 different ages (e.g., [30, 31, 32], with few exceptions ([33, 34], for example, collected up to 6  
41 hours of observational data longitudinally on the same child). Notwithstanding the limited  
42 amount of data on individual children, one common finding is that joint activity with peers is  
43 limited (they prefer to interact with adults, see e.g., [35]), rudimentary in their second year of  
44 life, and achieves a certain degree of complexity only around 2 years of age [36]. Though it  
45 should be acknowledged that some scholars have more recently pointed out that interactions  
46 between peers during the first year of life likely play an important role in a child socio-emotional  
47 development (see e.g., [37])  
48  
49

50 In this paper we provide a systematic investigation of the dynamic unfolding of  
51 spontaneous joint actions in two and four year olds while interacting with peers in preschools. By  
52 relying on methods mutated by primatology such as long-term focal observations across several  
53 days on the same subjects [38], we report on the frequency, duration and typology of joint  
54 actions young children engage in with their peers in a natural setting. Additionally, we consider  
55  
56  
57

1  
2  
3 the role played by several social variables in affecting the occurrence of signals that mark how  
4 they get into and out of joint actions. Among the joint actions we mostly focus on, there are  
5 different types of social play and conflict.  
6

7 Social play is probably the most common type of joint activity that young children  
8 engage in. Studies focused either on how much daily time is spent engaged in play or quantifying  
9 solitary vs. social play (e.g., [39]) or on the development of pretend play (e.g., [40]) as a proxy  
10 for symbolic cognitive abilities. One key interest of earlier observational studies was the ability  
11 of children to coordinate their actions and when they begin shifting from parallel play to social  
12 play to pretend play (see e.g., [34]). Other scholars have investigated how young children fight  
13 over toys (e.g., [41, 42, 43]), the frequency of conflicts in children (e.g., [44]), how relationships  
14 (friendship, kinship) affect the likelihood of conflict in young children (e.g., [45, 31] and the role  
15 played by individual differences and parental socialization on how conflicts are resolved (see,  
16 e.g., [45, 46]). Among the many things we know about conflict in young children, there is the  
17 fact that they mostly fight because of access to toys [47] or violation of personal space [48], the  
18 conflicts are short (between 15 and 24 seconds on average, [43, 48]), and the frequency and  
19 duration of conflict between friends and non-friends is analogous, though they seem less intense  
20 with friends [31]. Interestingly, conflicts between friends tended to have similar starts compared  
21 to non-friends but different endings, like getting resolved through mutual disengagement in the  
22 case of friends. Also, many studies did not find any effect of gender on the frequency and  
23 duration of conflict (see, e.g., [31, 49]).  
24

25 While play and fighting both entail sustained interactions between two or more  
26 individuals, play is cooperative, requires a certain level of coordination obtained through  
27 repeated signaling and an understanding of the ultimate shared goal, while fighting is  
28 competitive and does not require communicative signals to be sustained. Therefore, investigating  
29 how play and conflict are initiated, maintained and ended might provide additional information  
30 concerning how the degree of “jointness” of these social activities and their “cooperative” vs.  
31 “competitive” underlying motivations can be observed in children’s behaviour *in situ*. We will  
32 focus in particular on two types of social play: “play without object” and “play with object”. The  
33 first type will allow us to compare children’s behaviours to those recently reported for bonobos  
34 and chimpanzees ([50, 12] while the second might display human-specific features of play and  
35 joint actions. These two types of play will be specifically compared to “conflict”, for an  
36 ecologically relevant comparison between play and conflict. In addition to play and conflict, we  
37 intend to track joint action behaviors commonly observed in children, like grooming (e.g.,  
38 caressing someone’s hair) or socio-physical touch (e.g., holding hands) and the frequency of  
39 conversation with peers.  
40

41 This study combines exploratory and confirmatory components. From an exploratory  
42 perspective, we consider the frequency, duration and multi-party nature of several types of joint  
43 actions. We also consider the general sociability of two and four year old children, the  
44 observable strength of their social bonds and investigate several factors that might play a role in  
45 the dynamic unfolding of joint actions (gender, party size, siblings, strength of social  
46 relationships, etc.)  
47

48 From a confirmatory perspective, we formulate four general predictions based on the  
49 hypothesis that joint action coordination is contingent on the development of theory of mind,  
50 language and social norm orientation. First, we predict that four year olds will be more likely  
51 than two year olds to engage in social play. Second, we predict that number of social partners  
52 (partner quantity/social experience) and strength of relationship with the partners (level of  
53  
54  
55  
56  
57

friendship) will modulate the occurrence of entry and exit phases during joint actions, so that a higher number of social partners and stronger friendship should correlate with a reduced amount of entry and exit phases. Third, we predict that four year olds will be less likely than two year olds to engage in social conflicts. Fourth, we predict that there will be more entry and exit phases in play than in conflict, i.e. they will be more likely than in conflicts to elicit the gaze behaviour of the recipient, to produce communicative signals that convey their intentions to engage in the joint action and to communicate an intention to end the activity before doing so. Accordingly the behaviour during conflicts will likely be unilaterally directed, mostly insensitive to the attention and intentions of the recipient and its closing will most likely be achieved by a unilateral abandonment of the fight.

## 2. Methods

### 2.1 Subjects and Schools

Participants were 31 children attending four preschools in the same county in Southern California. There were 14 two-year-olds ( $M = 28.7$  months,  $SD = 2.75$ ; boys 6) and 17 four-year-olds ( $M = 51.3$  months,  $SD = 2.56$ ; boys 11), where age refers to the age on the first day of observation. The sample was 35% European American ( $N = 11$ ), 6% ( $N = 2$ ) Hispanic, 13% ( $N = 4$ ) multiethnic, and 45% ( $N = 14$ ) parents declined to provide race or ethnicity information in the demographic form. Seventeen of the 31 children had siblings. Children with siblings had a mean of 1.12 siblings ( $SD = 0.31$ ). 29.5% of focal children with siblings ( $N = 5$ ) were the oldest child among their siblings and 70.5% ( $N = 12$ ) were the youngest child among their siblings. There were no reported middle children.

All 4 preschools in which data was collected allowed children several hours of free play during the day. During free play, children would be interacting with several children of different ages. All preschools provided both outdoor and indoor spaces for free playing opportunities. During free play sessions (i.e., outside of specific curricula led by a teacher), teachers' interventions were usually limited to assisting in resolving conflicts if they escalated and monitoring children for safety.

### 2.2 Data Collection

We conducted focal observations on individual children between May 2017 and June 2018. Each focal subject was followed for a total of 20 hours (in 6/31 children, data collection had to be shortened to approximately 10 hours because the child left the school during the time longitudinal data was being collected or was not present for several data collection sessions). A total of 563 hours of video footage was collected. Usually, data collection on the same child would occur 2-3 times a week, leading to a focal follow on the same child lasting between 5 and 10 weeks total. No child was focal followed for more than 1 hour per day.

We filmed the children using digital High-Definition cameras (Canon Vixia HF G40 or Canon Vixia HF R800) and when possible relying on a directional shotgun microphone (Sennheiser MKE 600). We filmed the children always during free play time, to maximize comparability of the data across the four schools.



### 2.3 Video Coding Procedure

We coded the videos collected through ELAN 6.2 [52], a video annotation software designed to facilitate annotation of multimodal data. Given the shortened data collection for 6 out of the 31 subjects, we coded for joint actions in 10 hours of observation of each subject (or the amount available for each child, in some cases slightly less). Specifically, we coded for the frequency and duration of 6 different joint action types (see Table 1 for definitions). We excluded any joint action lasting less than 3 seconds. We also identified the interactional partner of each focal subject for each joint action event. For 5 out of the 6 joint action types (excluding conversation), we further coded the occurrence of an entry phase, a main phase and an exit phase drawing on the same coding method as in [12], to ease comparability of the results to what observed in non-human primates. The coding of approximately 310 hours of data led to the identification of 4019 total joint actions. Moreover, we further coded a subset of 5 out of 6 joint action types (excluding Conversation) for the occurrence of entry, main and exit phases and their durations. Specifically, we further coded 585 joint actions, amounting to approximately 27% of the joint action events available within that subset of joint action types.

**Table 1.** Definition of Joint Action Types identified in the dataset.

JOINT ACTION TYPE	DEFINITION
Play with Object	Focal child will <i>jointly/take turns</i> with another child engaging in conversation, singing, dancing, chasing, hiding, racing, tagging, pretend play, or rough-housing while interacting with the same objects. There must be clear communication between the focal child and the other child/children (i.e. they are not engaged in parallel play with the same object)
Play without Object	Focal child will <i>jointly/take turns</i> with another child engaging in conversation, singing, dancing, chasing, hiding, racing, tagging, pretend play, or rough-housing. There must be clear communication between the focal child and the other child/children (i.e. they are not engaged in parallel play)
Social Physical Touch	Focal child makes affiliative physical contact with another child (e.g., holding hands, high fiving, tapping shoulder to get attention, etc.)
Grooming	Focal child's hair is touched (e.g, stroking, brushing hair, etc.) and/or moved around (e.g., sculpting a hairstyle) with either hands or with an object (e.g., brush, comb, fork, etc.) OR the focal child touches and moves another child's hair

<p>Conversation</p>	<p>Focal child and another child are reciprocally engaged in vocal communication. This will be coded <i>only</i> when conversation does <i>not</i> occur during other joint-activities here identified and when at least one verbal utterance by one child is responded to by the other child.</p> <p>If Focal and classmate converse while engaged in any other joint activity coded for this study (e.g. play, conflict, grooming), this is coded with the label of the larger activity they are part of (e.g. Play with object) and not as Conversation</p>
<p>Conflict</p>	<p>When the focal child and another child disengage from cooperative/neutral behavior and cannot reconcile opposing goals or interest in a toy without protest or physical interruption. Often accompanied by a change in register <u>in both children</u> (e.g., using more aggressive or diplomatic language to resolve the conflict) and a change in body language (e.g., nonverbal cues that show the child is upset or sad. Some examples include: crossing arms, hanging their head, scowling, covering their face).</p> <p>The behavior must be exhibited by both children to count as “Conflict”. If one child ignores the other, it is not considered as a joint action.</p>

Following [12] we defined an entry phase as “the process by which partners recruit each other, via mutual gaze exchanges and intentional communicative signals (for intentionality criteria, see [53]) to determine the type of the activity, negotiate its temporal and spatial properties and potentially establish joint commitment. Because to enter a joint action, both partners need to mutually agree to it, entries are always achieved through (at minimum) the exchange of mutual gaze (i.e., both partners simultaneously look at each other’s face) but can also include other signals such as gestures or vocalizations (e.g. greetings). If partners did not engage in mutual gaze exchanges and did not produce any other communicative signal before engaging in the activity itself, the entry phase was coded as absent.” (p. 17) In other words, the occurrence of mutual gaze alone or just greetings between the children without mutual gaze was sufficient for that joint action to be coded as having an entry. We defined an exit phase as “the process by which partners [...] take leave of each other. They may express intentions to end a joint action via intentional communicative signals or other specific behaviors before walking away. [...] If partners did not engage in gaze exchanges and did not produce any other communicative signal before leaving their partner, the exit phase was coded as absent.” [12 p. 17] The main body phase consisted in the movements typical of the joint activity, such as chasing each other for play or holding hands for socio-physical touch. The main body phase could occur without the occurrence of an entry or an exit phase.

Before proceeding with coding, all coders had to pass a reliability threshold. After some initial training on definitions, coders were asked to code all instances of joint action first and later phases on the same two 30-minute videos from different children that were part of the data



not meant to be coded for joint actions. A lab manager had previously coded these videos as well and that performance was taken as the standard. The research assistants were selected for further coding of the behavioral data if they passed a threshold of 70% agreement with the lab manager in terms of identifying the presence of a joint action within 5 seconds of when it had been identified by the lab manager and whether they had labeled it as the same type of joint action. 5 out of 6 coders passed the threshold (ranging in performance between 71.2% agreement,  $K=0.63$  to 81% agreement,  $K=0.75$ ) and then proceeded to code 310 hours of video data. The coder excluded had obtained a score of 54.6%,  $K=0.42$ .

## 2.4 Statistical Analysis

This study considered 6 joint action types resulting in 4019 total joint actions: conversation ( $N = 1859$ , 46% of dataset), play with object ( $N = 1229$ , 30% of dataset), play without object ( $N = 508$ , 13% of dataset), social physical touch ( $N = 223$ , 5% of dataset), conflict ( $N=219$  5%), and grooming ( $N = 16$ , <1% of dataset). A randomly selected subset of these joint actions was coded for entry/main body/exit phases resulting in 585 fully annotated activities which did not include conversation: play with object ( $N = 338$ , 54% of subset), play without object ( $N = 144$ , 23% of subset), social physical touch ( $N = 65$ , 11% of subset), conflict ( $N = 62$ , 11%), and grooming ( $N = 6$ , 1% of subset). The subset coded for phases included both dyadic and multiparty joint actions.

Statistics were calculated with python v3.8 using the *GLM* function in the statsmodels package v12 and with R v4.1.2 using the *glmer* in the package lme4 v1.1.12. We ran ordinary least squares regression (OLS) to examine sources of variation in (a) the rates that children engage in joint activities and (b) the relative time children engage in joint activities. For these two statistical analyses, data were aggregated per child and activity type, so we did not need to account for multiple comparisons. We ran a linear mixed model (LMM) to examine sources of variation in (c) main body phase durations. We ran generalized linear mixed models (GLMMs) with binomial error and the logit link function to examine sources of variation in (d) the presence of entry and exit phases. For all models we considered age group, gender, and the activity type (using play with object as our reference treatment) as independent variables. For models (c) and (d) focal child ID and partner ID were included as random effects. For models (c) and (d) we additionally considered the activity group size (how many participants were engaged in a joint action) (min=2, max=7), whether the focal child had siblings (True or False), the relative dyad co-occurrence (RDC) and number of recurring partners (NRP). RDC (min=0.002, max=0.490) is a proxy for closeness of relationship/friendship and is defined as

$$RDC(p_i, p_j) = \frac{JA(p_i, p_j)}{JA(p_i)}$$

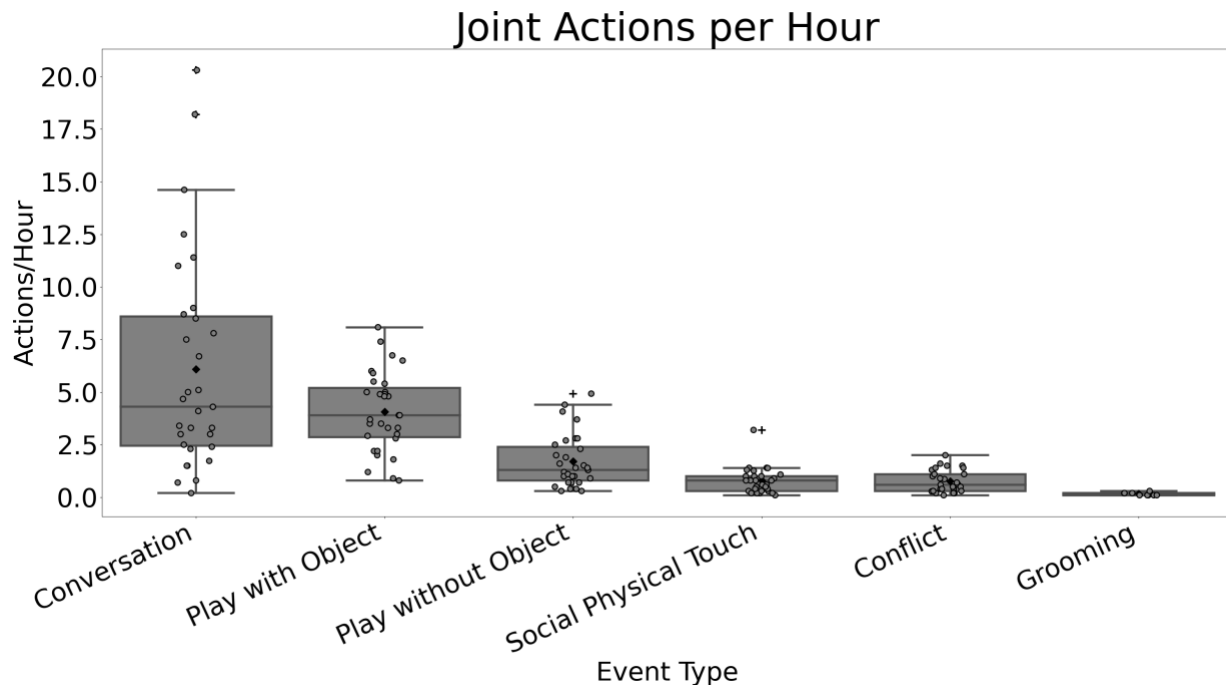
Here,  $p_i$  is a participant.  $JA(p_i)$  is the number of joint actions  $p_i$  was involved in.  $JA(p_i, p_j)$  is the number of joint actions both  $p_i$  and  $p_j$  were involved in. When there were more than two children in a joint action, we used the maximum RDC score between the focal child and other children. RDC represents how frequently a child is interacting with the focal child out of the total joint actions that the focal child participates in. NRP (min=8, max=24) is a proxy for social experience and is defined as the number of distinct partners a focal child interacts with more than once (partner quantity). We computed RDC and NRP using a subset of the dataset that includes all 3835 positive joint actions, i.e., all conversation, play with object, social physical touch, and grooming but no conflict.

### 3. Results

#### 3.1 Descriptive Statistics

##### 3.1.1 Social Partners

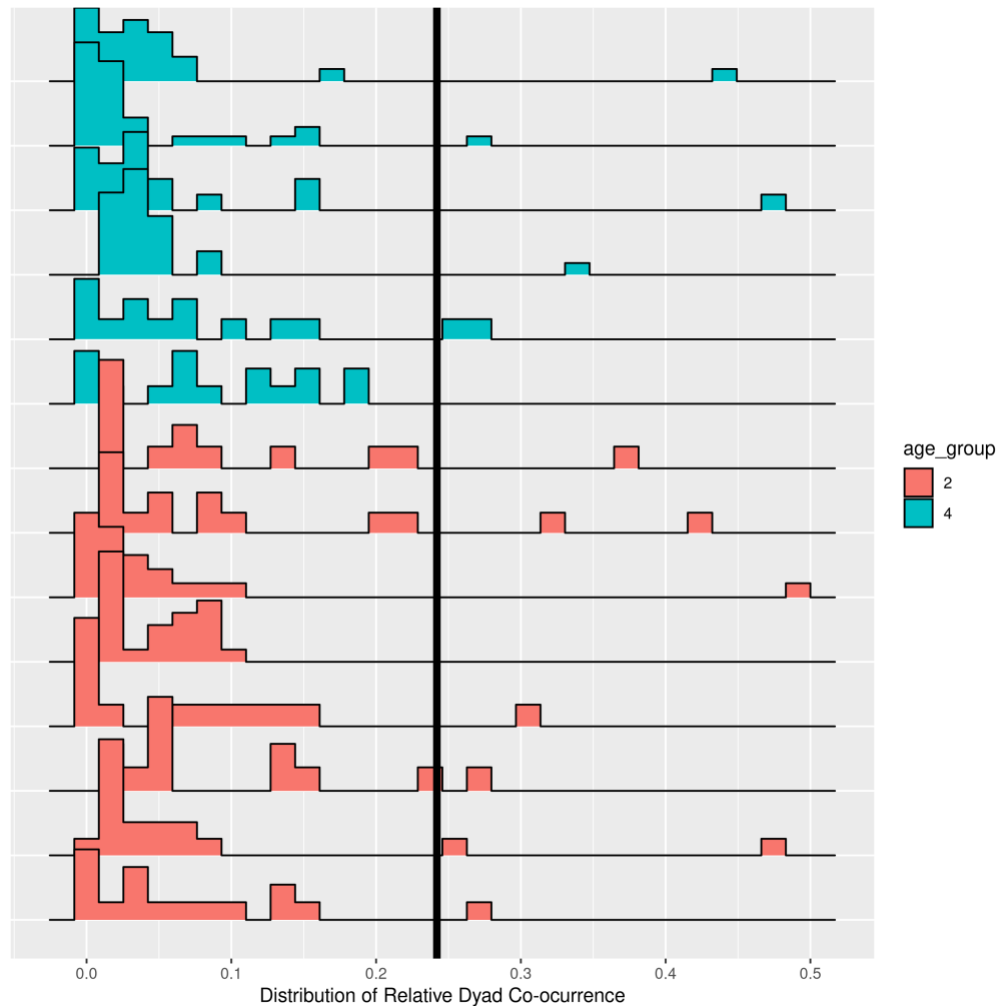
First, we report on the type of sociality we have identified in the children participating in this study. On average, within the 10 hours of data coded for joint actions, children engaged with 14.29 repeated partners, i.e. social partners they engaged with more than once during that time. The range extending from a minimum of 8 social partners to a maximum of 24.



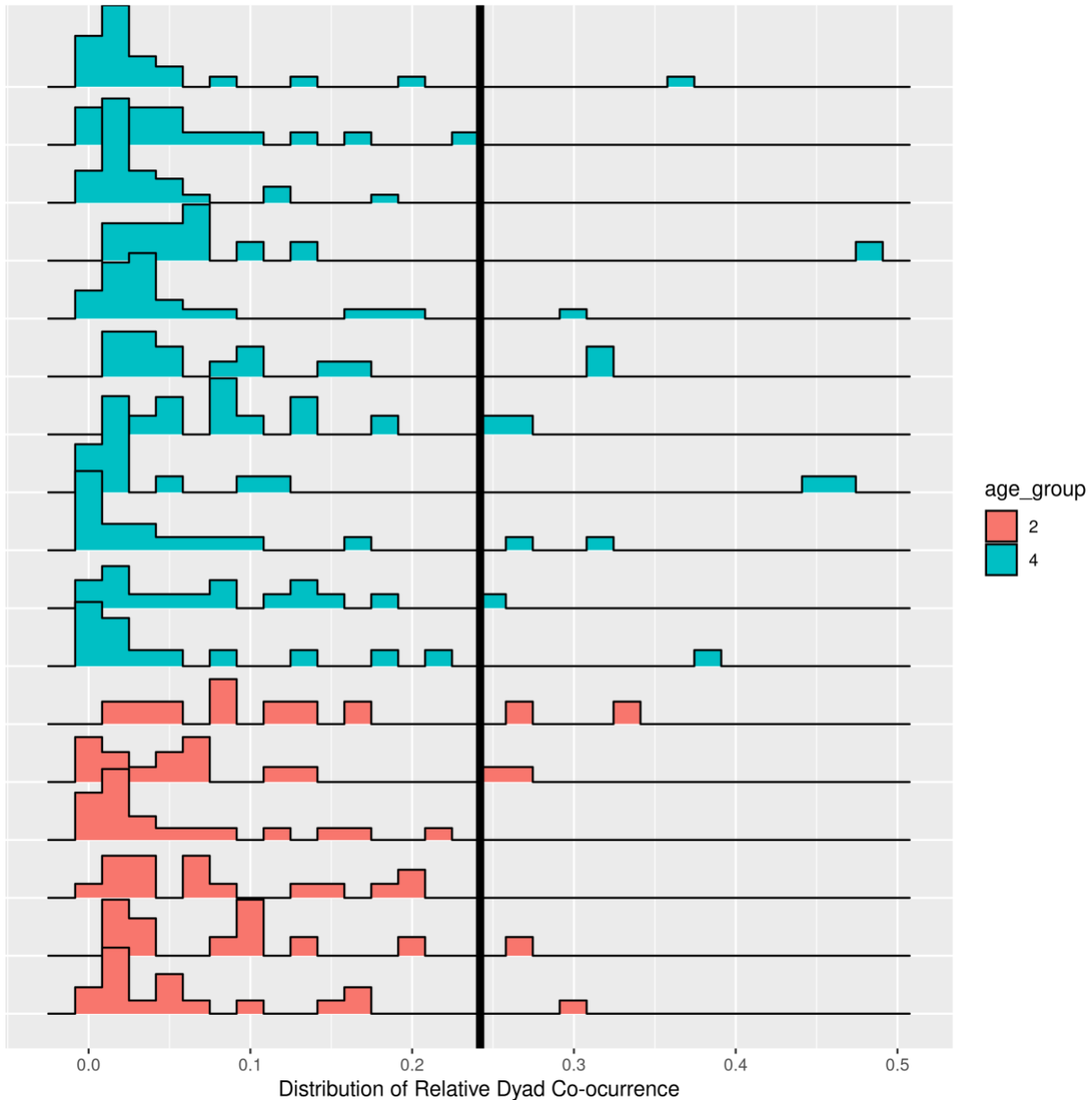
**Figure 1:** Distribution of the number of recurring joint action partners for each focal child (NRP score). Mean=14.29 repeat partners, SD= 4.07

When we look at the frequency focal subjects interacted with each of these social partners, a clear pattern emerges: the vast majority of children have several social partners they interact rarely with and one to two individuals they interact much more often than anyone else. This is true for both boys and girls and for both two and four year olds. We can think of these individuals as focal subjects' closest friends (Hinde, 1976). Specifically, 25 out of 31 children had at least one partner with whom they interacted in more than 25% of the total joint actions they engaged in, which represents the equivalent of the mean dyad co-occurrence in terms of joint action +2 standard deviations (see Figures 2 and 3). Boys mean RDC was  $0.074 \pm 0.088$  with a minimum of 0.003 and a maximum of 0.485. 13 of 17 focal boys had at least one partner above the outlier threshold. Focal boys had a mean of  $1.6 \pm 0.518$  partners above the outlier threshold. Girls mean RDC is  $0.067 \pm 0.088$  with a minimum of 0.002 and a maximum of 0.494. 12 of 14 focal girls had at least one partner above the outlier threshold. Focal girls had a mean of  $1.25 \pm 0.452$  partners above the outlier threshold.

Overall mean RDC is  $0.070 \pm 0.08$  with a minimum of 0.002 and a maximum of 0.493. 25 of 31 focal children had at least one partner above the outlier threshold. Focal children had  $1.36 \pm 0.489$  partners above the outlier threshold.



**Figure 2:** Ridgeplot showing distributions of relative dyad co-occurrence for each *boy*. The black vertical line indicates an outlier threshold ( $\text{mean} + 2\text{SD} = 0.242$ ) where mean and SD are computed over RDC for all focal children after removing outliers using the outliers R package. The individuals on the right of the black vertical line are the close friends.



**Figure 3:** Ridgeplot showing distributions of relative dyad co-occurrence for each *girl*. The black vertical line indicates an outlier threshold ( $\text{mean} + 2\text{SD} = 0.242$ ) where mean and SD are computed over RDC for all focal children after removing outliers using the outliers R package. The individuals on the right of the black vertical line are the close friends.

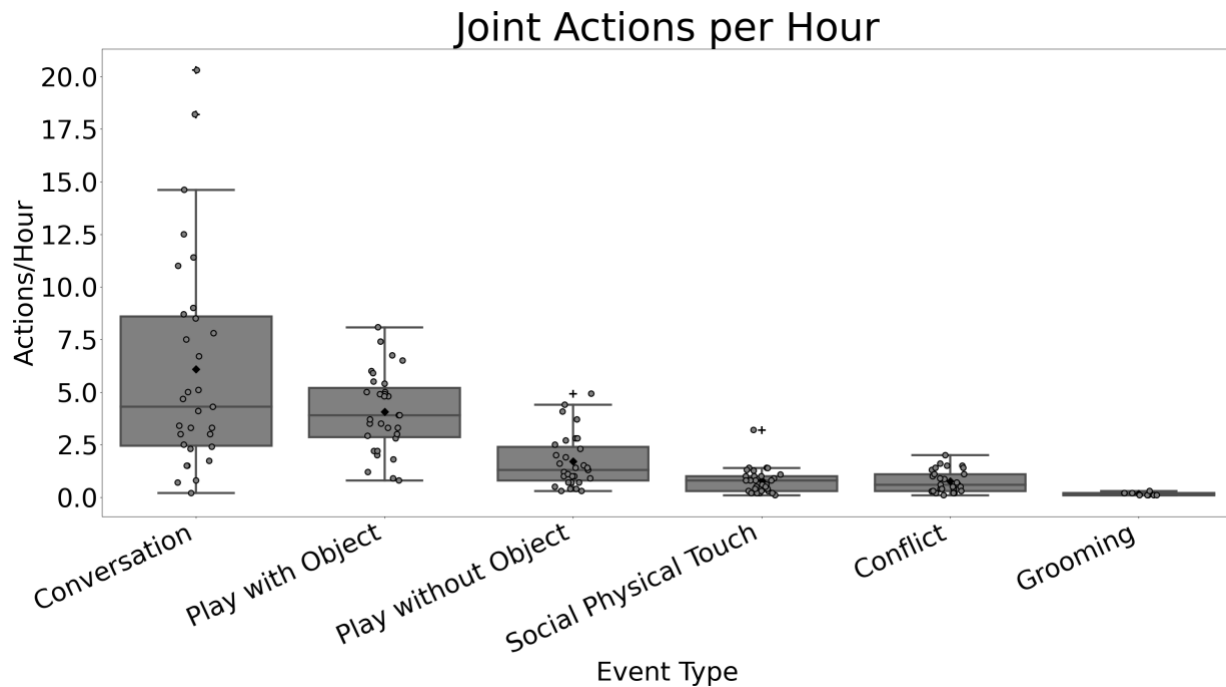
We then considered the frequency with which the joint actions performed by two and four year olds were dyadic, vs. multi-party (see Table 2). Overall, 82.7% of joint actions were performed dyadically with little difference between two and four year olds (84.6% vs. 81.6%). Notably while almost all conflicts and social-physical touch were performed dyadically, a quarter of play events occurred with more than 2 participants at a time.

**Table 2.** Proportion of Joint Actions Performed Dyadically

	2 years old		4 years old		total
	Female	Male	Female	Male	
Conversation	88.2	88.9	83.8	86.5	86.3
Play with Object	83.1	75.7	75.4	79.1	78.5
Play without Object	71.9	77.6	64.9	69.3	70.5
Social Physical Touch	87.0	92.7	91.4	91.0	90.5
Grooming	100.0	100.0	71.4	66.7	75.0
Conflict	94.8	97.6	96.4	98.6	96.8
	2 years old		4 years old		total
	84.6		81.6		82.7

### 3.1.2 Joint Actions

From the 4019 joint actions, we observe that on average, individuals engaged in 13.36 ( $\pm 7.23$ ) joint actions per hour



**Figure 4:** Joint actions per hour for each activity type. Shown are individual rates (grey dots), population means (filled diamonds), medians (horizontal lines), quartiles (boxes), percentiles (5% and 95%, vertical lines) and outliers (black crosses).

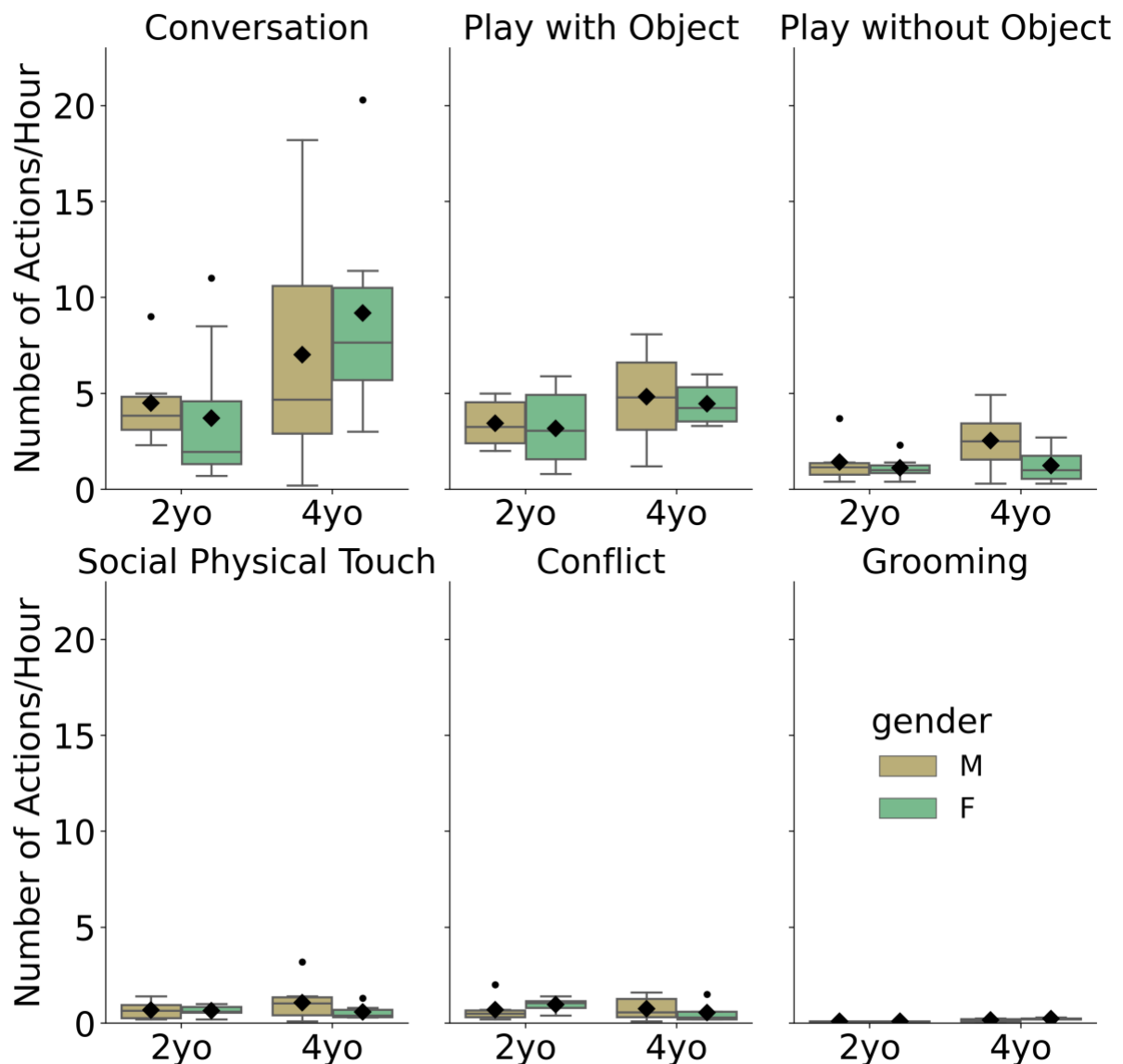
### 3.2 Statistical Models

### 3.2.1 Effect of age and gender on likelihood of engaging in different types of joint actions

To test our first prediction (older children will be more likely to engage in social play than younger ones), we examined whether children of different ages and genders engage in joint actions at different rates (defined as the number of joint actions per hour) and whether these rates differed by activity type. We aggregated over the dataset of 4019 joint actions to determine the number of joint actions per hour for each child per activity type, resulting in 161 rates. Using OLS regression, we tested sources of variation in these rates. The results of the model indicate the predictors explained 51.8% of the variance ( $pseudo-R^2 = .558$ ,  $chi^2 = 86.6$ ). In line with our first prediction, we found a significant effect of age group ( $\beta \pm SE = 0.177 \pm 0.061$ ,  $z = 2.918$ ,  $p = .001$ ), with older children engaging in more joint activities. Moreover, we found children engage in different joint activity types at different rates compared to play with object: conversation ( $0.630 \pm 0.191$ ,  $z = 3.300$ ,  $p = .001$ ), play without object ( $-0.729 \pm 0.191$ ,  $z = -3.814$ ,  $p < .001$ ), social physical touch ( $-1.013 \pm 0.194$ ,  $z = -5.210$ ,  $p < .001$ ), conflict ( $1.008 \pm 0.194$ ,  $z = -5.187$ ,  $p < 0.001$ ), grooming ( $-1.261 \pm 0.274$ ,  $z = -4.598$ ,  $p < .001$ ). There was no significant effect of gender ( $0.042 \pm 0.122$ ,  $z = 0.342$ ,  $p = .691$ ). This means that children engage in more conversations per hour than in joint actions involving play with objects but fewer joint actions involving play without object, social physical touch, grooming or conflict compared to play with object.

Figure 5 shows the frequency of joint actions broken down by age group, gender, and activity type.





**Figure 5:** Joint actions per hour in relation to age, gender, and activity type. Shown are population means (filled diamonds), medians (horizontal lines), quartiles (boxes), percentiles (5% and 95%, vertical lines) and outliers (filled dots).

### 3.2.2 Effect of age group, gender, and activity type on relative time engaged in joint actions

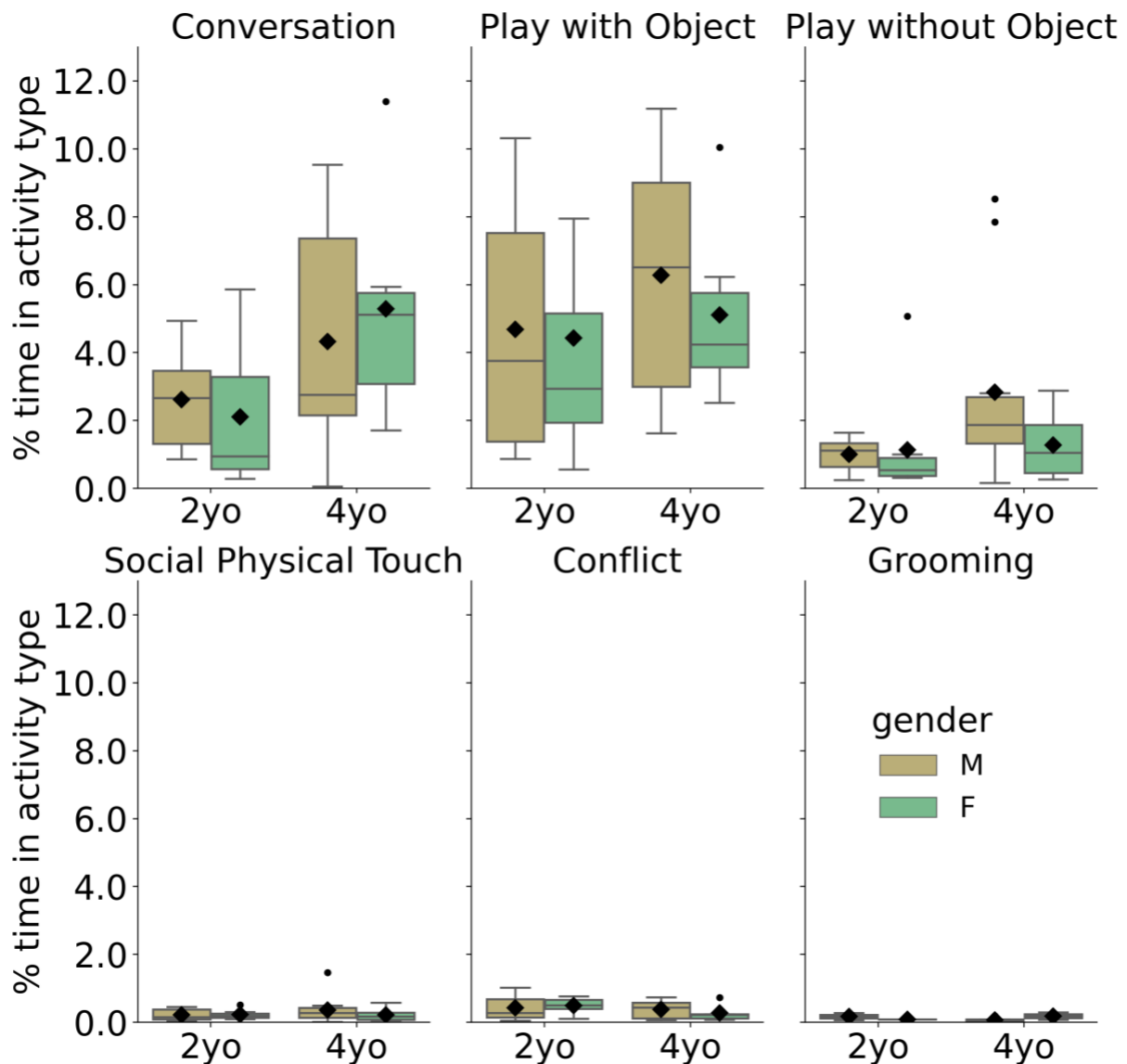
We further tested whether children of different ages and genders spend different amounts of time engaged in joint actions (as measured by the sum of joint action durations divided by total coded observation time) and whether these durations differ by activity type. We aggregated over the dataset of 4019 joint actions to determine the total duration of joint actions relative to observation duration per child per activity type, resulting in 161 relative durations. Using OLS regression, we tested sources of variation in these relative durations. The results of the model indicate the predictors explained 50.2% of the variance ( $pseudo-R^2 = .558$ ,  $chi^2 = 86.7$ ). We found a significant effect of age group ( $\beta \pm SE = 0.150 \pm 0.061$ ,  $z = 2.473$ ,  $p = .031$ ), with older

children engaging in joint activities longer than younger children. Moreover, we found children engage in different joint activity types at different rates compared to play with object: grooming ( $-1.800 \pm 0.274$ ,  $z = -6.559$ ,  $p < .001$ ), social physical touch ( $-1.701 \pm 0.194$ ,  $z = -8.747$ ,  $p < .001$ ), conflict ( $-1.646 \pm 0.195$ ,  $z = -8.462$ ,  $p < .001$ ), play without object ( $-1.200 \pm 0.191$ ,  $z = -6.288$ ,  $p < .001$ ), and conversation ( $-0.565 \pm 0.191$ ,  $z = -2.953$ ,  $p = .003$ ). Despite engaging in 50% more conversations, children spend more time playing with objects (see Table 3). There was no significant effect of gender ( $-0.092 \pm 0.122$ ,  $z = 0.755$ ,  $p = 0.450$ ).

From the 4019 joint actions, we observe that the average joint action lasted 27.52 ( $\pm 22.80$ ) seconds (including entry and exit phases, as well as the main body phase) and individuals spend on average 11.28% ( $\pm 6.57\%$ ) of their time engaged in joint activities (See Table 3 and Figure 6 for detailed results broken down by age group, gender, and activity type).

**Table 3:** Average duration (seconds) of joint actions in relation to age group, gender, and activity type.

	2 years old				4 years old				total	
	Female		Male		Female		Male		Mean	SD
	Mean	SD	Mean	SD	Mean	SD	Mean	SD		
Conversation	19.45	5.36	22.20	15.25	21.04	4.63	21.61	8.06	21.06	8.47
Play with Object	49.03	30.63	44.23	34.35	43.62	29.70	46.86	17.70	46.29	25.91
Play without Object	32.91	28.65	28.24	8.38	47.87	50.76	37.74	23.64	36.61	29.37
Social Physical Touch	12.63	5.77	11.15	4.59	12.58	7.43	10.18	4.33	11.47	5.29
Conflict	18.16	5.67	22.55	19.44	20.08	12.36	21.18	11.49	20.44	11.93
Grooming	30	NA	61	53.74	26.66	11.93	12.87	9.47	28.35	26.96



**Figure 6:** Percent of time spent engaged in joint action in relation to age, gender, and activity type. Shown are population means (filled diamonds), medians (horizontal lines), quartiles (boxes), percentiles (5% and 95%, vertical lines) and outliners (filled dots).

### 3.2.3 Durations and rates of joint action entries and exits

Having reported on frequency and duration of joint actions and on the number of social partners and different types of relationships between children, we can now consider the way they start and the way they end. Two year olds and four year olds signaled entries at similar rates: 66.3% and 69.0% of joint actions respectively. The average duration of an entry phase was 2.97 seconds for two year olds and 2.93 seconds for four year olds.

Two year olds and four year olds signaled exits at similar rates: 69.3% and 75.5% of joint actions respectively. The average duration of an exit phase was 2.56 seconds for two year olds and 2.27 seconds for four year olds.

**Table 4:** Proportion of joint actions that contain entry by age, gender and activity type

	Entry				total
	2 years old		4 years old		
	Female	Male	Female	Male	
Play with Object	67.7	67.6	76.0	74.6	71.9
Play without Object	68.7	84.0	54.1	71.3	69.8
Social Physical Touch	54.7	48.6	77.7	56.1	58.7
Conflict	56.2	81.6	85.0	78.5	74.1
Grooming	100.0	NA	33.3	50.0	50.0
	2 years old		4 years old		total
	66.3		69.0		67.8

**Table 5:** Proportion of joint actions that contain exit signals by age, gender and activity type

	Exit				total
	2 years old		4 years old		
	Female	Male	Female	Male	
Play with Object	75.4	75.7	64.2	69.6	71.2
Play without Object	54.2	88.9	79.4	31.1	74.9
Social Physical Touch	73.8	90.2	75.0	84.3	81.1
Conflict	44.7	56.6	50.0	74.0	58.2
Grooming	100.0	NA	100.0	100.0	100.0
	2 years old		4 years old		total
	69.3		75.5		72.7

### 3.2.4 Factors affecting the presence of entries and exits

Next, to test our second prediction (the number of social partners and strength of social bond will affect the occurrence of entry and exit phases), we examined whether the occurrences of entry and exit phases differed across age groups, genders, activity types, activity group sizes, NRP, RCD, and whether the focal child had siblings. To test this, we used two GLMMs (one for entries and one for exits) with binomial error and logit link function for these sources of variation in entries and exits. We excluded instances of grooming due to the limited number of observations of this activity type.

1  
2  
3  
4  
5  
6  
7 Contrary to our prediction, we found no significant effects on the presence of entries (See Table  
8 6). We also found no effects on the presence of exits (See Table 7).  
9

10 **Table 6:** GLMM estimates entries. Estimates of Joint Action types are with respect to Play with  
11 Object.  
12

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	0.134	0.636	0.211	0.833
Conflict	-0.049	0.317	-0.154	0.878
Play without Object	-0.236	0.231	-1.022	0.307
Social Physical Touch	-0.536	0.300	-1.784	0.074
age group	0.123	0.142	0.863	0.388
gender[M]	0.104	0.267	0.390	0.697
RDC	0.679	0.824	0.824	0.410
NRP	0.001	0.035	0.021	0.983
has sibling	-0.070	0.264	-0.266	0.790
group size	0.145	0.130	1.110	0.267

**Table 7:** GLMM estimates exits. Estimates of Joint Action types are with respect to Play with Object.

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	1.024	0.617	1.660	0.097
Conflict	-0.404	0.305	-1.324	0.185
Play without Object	0.154	0.241	0.642	0.521
Social Physical Touch	0.282	0.329	0.858	0.391
age group	0.022	0.138	0.160	0.873
gender[M]	0.452	0.255	1.771	0.077
RDC	-0.273	0.814	-0.335	0.738
NRP	-0.038	0.033	-1.146	0.252
has sibling	0.211	0.254	0.831	0.406
size	0.036	0.127	0.285	0.776

### 3.2.5 Factors affecting main body phase duration

Next, we tested whether main body phase durations differed across age groups, genders, activity types, activity group sizes, NRP, RDC, and whether the focal child had siblings. We used an LMM to test sources of variation in main phase duration. If there were multiple other children involved in a joint activity, we used the greatest RDC score between the focal child and other children. The results of the model indicate the predictors explained 17% of the variance (pseudo- $R^2=0.172$ ,  $\chi^2=438$ ).

We found a significant effect of activity group size on main body phase duration ( $\beta \pm SE = 0.339 \pm 0.060$ ,  $t = 5.625$ ,  $p < 0.001$ ), meaning that children engaged in longer joint actions when there were more than 2 partners involved. We found a significant effect of gender: compared to girls, boys engaged in shorter joint actions ( $-0.254 \pm 0.117$ ,  $t = -2.169$ ,  $p = 0.043$ ). We also found a significant effect of NRP ( $-0.036 \pm 0.016$ ,  $t = -2.306$ ,  $p = .032$ ), meaning that children who had more repeated partners engaged in shorter joint actions. Additionally, we found a significant effect of activity type. Compared to play with object, social physical touch was associated with shorter joint actions ( $-0.926 \pm 0.144$ ,  $t = -6.451$ ,  $p < .001$ ) as was play without object ( $-0.239 \pm 0.106$ ,  $t = -2.254$ ,  $p = .025$ ).



### 3.2.6 Factors affecting the length of main body phases.

Next, we tested whether exit phases were less common when the duration of the main body phase of a joint action was longer. We used a GLMM with binomial error and logit link function to test sources of variation. We found a significant effect of the log duration of the main body phase on the presence of exits ( $\beta \pm SE = -0.205 \pm 0.094$ ,  $z = -2.189$ ,  $p = .029$ ). This means that when joint actions have longer main body phases focal children are less likely to coordinate an exit phase when disengaging from interactions.

### 3.2.7 Factors affecting Conflict

Focusing specifically on conflict and to address our third prediction (older children will be less likely to engage in conflict compared to younger ones), we investigated age group, gender, activity group size, NRP, greatest RDC, and whether the focal child had siblings as sources of variation between instances of conflicts and other cooperative joint activity types. To test this, we used a GLMM with binomial error and a logit link function. Our dependent variable was whether or not an interaction was a conflict ( $y=1$ ) or not ( $y=0$ ).

In line with our third prediction, we found a significant effect of age group in discriminating between conflicts and non-conflicts ( $\beta \pm SE = -0.433 \pm 0.192$ ,  $z = -2.260$ ,  $p = .024$ ) meaning that four year olds were less likely to engage in conflicts than 2 year olds. We also found a significant effect of NRP ( $0.087 \pm 0.041$ ,  $z = 2.124$ ,  $p = .034$ ) meaning that children who overall interacted with more children were more likely to engage in conflicts. Additionally, we found a significant effect of greatest RDC ( $-7.688 \pm 2.150$ ,  $z = -3.417$ ,  $p = .001$ ), meaning that children are less likely to engage in conflicts with those they interact more frequently with (their friends).

To address our fourth prediction (that conflicts would have less entry and exit phases than other cooperative joint actions), we examined whether the occurrences of entry and exit phases differed across age groups, genders, and whether an activity was a conflict or non-conflict. To test this, we used two GLMMs (one for entries and one for exits) with binomial error and logit link function for these sources of variation in entries and exits. For additional clarity, unlike the model testing hypothesis 3, in which conflict was a predicted variable, in these models testing hypothesis 4, conflict is treated as a binary predictor variable, e.g., conflict=1 or conflict=0. Contrary to our fourth prediction, we found no significant effect of conflict on likelihood of a joint action event to have an entry ( $0.017 \pm 0.300$ ,  $z = 0.055$ ,  $p = .956$ ), while controlling for age, gender and focal child and partner ID. Conflict does not have a clear effect on exit either ( $-0.529 \pm 0.288$ ,  $z = -1.832$ ,  $p = .067$ ). Finally, gender does not appear to affect likelihood of conflict in preschoolers ( $0.457 \pm 0.256$ ,  $z = 1.784$ ,  $p = .074$ ).

## 4. Discussion

Towards investigating the development of the human interaction engine, this paper has focused on the factors affecting how and how often 2- and 4-year-children engage in joint actions with peers. Combining the results across the two ages, we found that overall young children engage in joint actions frequently (on average 13 times per hour), often with more than one partner (in almost 20% of cases) and with several different social partners. These joint actions are often brief (on average 28 seconds), and contrary to expectations they do not have

1  
2  
3 entry and exit phases as frequently as reported in great apes [12]. They engage mostly in  
4 conversation and social play with objects and these two joint actions are also the ones with the  
5 longest duration. This produces several opportunities for children to practice how to enter and  
6 exit social situations because of many fast-paced interactions with a diverse pool of social  
7 partners. While the interaction engine hypothesis does not make any specific claim about the  
8 importance of number of social partners towards the development of human-like interactional  
9 ability, we would like to suggest that it might actually be an important variable. Recent  
10 anthropological work on tool making has shown for example that while hunter-gatherer men like  
11 Hadza and Ache might observe at least 300 other create tools in a life time, male chimpanzees  
12 are likely to observe no more than 20 other males because of their xenophobia (see [54]). The  
13 claim there was that diversity in models leads to improved social learning opportunities,  
14 innovation and cumulative culture. Similarly, it is likely that young human children interact with  
15 more social partners by age 4 than the average chimpanzee in a lifetime (though further  
16 empirical evidence is needed). This environmental exposure to diverse partners, and the  
17 difficulty of navigating the range of familiarity and social relationships children might have with  
18 those partners, would lead to the need to carefully coordinate signaling and monitoring each  
19 other's behavior.  
20  
21  
22

23 When we look at our original predictions concerning the development of joint action  
24 abilities in preschoolers, we find that in line with our first prediction, four year olds engage in  
25 cooperative joint actions more often than two year olds. An increased successful engagement in  
26 joint actions is in line with the expectation that older children would be able to better predict the  
27 behavior of their partners, align their conduct with them and signal their intention to start or end  
28 a joint activity. Cognitive development likely facilitates the ability of parsing others' behavior  
29 and coordinating with them. Surprisingly however, and in contrast with our second prediction,  
30 social experience and strength of social relationship do not seem to play any role in the  
31 sequential unfolding of joint actions (both cooperative and competitive ones), in particular in  
32 terms of affecting the likelihood of having an entry or exit phase.  
33

34 In line with our third prediction, two year olds fight more than four year olds and conflict  
35 is affected by the number of repeated social partners (the more you have, more likely you are to  
36 fight) and children are less likely to fight with close friends than with other children. More social  
37 partners often means more individuals that the child rarely interacts with. On the other hand,  
38 repeated interactions with the same social partner can lead to both an ability to better read the  
39 conduct of the other and predict what they are going to do next and an increased motivation to  
40 care for the social partner and therefore an interest in limiting the chances of conflict. Almost all  
41 two and four year olds have one or two close friends they interact often with and the number of  
42 close friends does not seem to be affected by gender. In contrast with our fourth prediction we  
43 find that conflicts is as likely as cooperative joint actions to have entry and exit phases,  
44 practically disconfirming that unilateral disengagement is more likely in conflicts compared to  
45 play situations.  
46  
47  
48

49 Concerning frequency and duration of conflicts, Hay [55] reports conflicts occurring on  
50 average 8 times per hour. Dawe [42] reports that conflicts are short, approximately 24 seconds  
51 on average and Houseman [56] reports even shorter average duration, approximately 12 seconds.  
52 While the duration of conflicts in our dataset (20.4 seconds) is analogous to what reported by  
53 Dawe, the frequency of conflict is significantly smaller, with less than 1 conflict per hour  
54 identified. The main reason for this difference likely comes from the way we identify conflict. In  
55  
56  
57

1  
2  
3 our study only conflict that is ratified by a recipient and that leads to joint actions would count.  
4 As such, situations in which one child might yell at the other or even push the other to the ground  
5 without getting a response would not be considered in our analyses.  
6

7 Similarly to what has been reported in earlier observational studies on play and conflict in  
8 toddlers, gender appears to have little to no effect on joint action frequencies and durations but it  
9 has an effect on duration of the main phase, with girls engaging in joint actions with longer main  
10 phases compared to boys. Noticeably, joint actions with longer main phases are less likely to  
11 show exits and main body phases tend to be longer when more children are involved. In our  
12 dataset about 25% of play activities involve more than two children, once again confirming the  
13 hyper-sociality of human children in line with the interaction engine hypothesis.  
14

15 The most noticeable finding here is the fact that only two thirds of joint actions have  
16 entry and exit signals (66-69%), similar to what was reported for chimpanzees but noticeable less  
17 than what was reported for bonobos (more than 90%). The fact that social relationship does not  
18 appear to modulate the occurrence of entry or exit signals in young children is also analogous to  
19 what observed in chimpanzees and yet different from what was claimed both for bonobos [12]  
20 and for human adults [13], although precise quantitative studies in naturalistic settings are  
21 lacking for adult humans. All of the above suggests that at least in young children, social  
22 relationships and the specific motivation to engage in joint actions, be it cooperative like in play  
23 or competitive like in conflict, might not be particularly important for the process through which  
24 joint actions are accomplished. It is also possible that for young children, engaging in play or  
25 conversation with peers entails less sharedness of goals and joint commitment than noted in  
26 experimental settings and as such more similar to what observes in chimpanzees. Note also that  
27 it is possible that in the preschool context, children were familiar enough with each other to  
28 make the occurrence of entries and exits much less necessary when compared with interactions  
29 with peers who are real strangers (i.e. never met before). This would suggest that we are  
30 underestimating the effect of social relationship because the range we are measuring is not wide  
31 enough. Further empirical studies will need to address this potential alternative explanation.  
32

33 Surprisingly the proportion of joint actions with entry and exit phases are practically  
34 identical across the two age groups. This proportion potentially suggest that the occurrence of  
35 such phases and the signaling associated with them might be less normative than originally  
36 predicted (a norm that is followed only 2/3 of the times does not have very normative power). It  
37 is likely that other specific situational factors can account for when entries and exits would  
38 actually occur, and those factors could likely be when the two individuals had most recently  
39 interacted and the valence of that interaction. Such relational factors have for example been  
40 shown to be relevant for the occurrence of attempts at food transfers in orangutans (see [57]) and  
41 the selection of which signals to use [58]. It is also possible that young children's signal  
42 production is affected by their reading of the social situation as a "continuing state of incipient  
43 talk" [59], i.e., social situations in which interactional lapses can occur between participants but  
44 that do not require renewed exchanges of greetings for example because participants remain  
45 perceptually accessible to each other (e.g., two people sitting side by side on a plane or watching  
46 a show). In this sense, similarly to what has been described for adult humans, if child A and child  
47 B have been playing together for a while but then moved to solo activities for a few minutes  
48 before interacting again, they might be thinking of the solo activities as interruptions of a larger  
49 social play activity with that specific partner. Given that our focal data was not collected always  
50 at the very beginning of the school day and we do not have access to all social interactions  
51 between a focal child and all other partners within a day, this remains a working hypothesis.  
52  
53  
54  
55  
56  
57

Overall, we believe we have provided a fine-grained picture of the extent to which two and four year olds spontaneously engage in joint activities with peers in preschools and the several factors that appear to affect joint action as a process. While the number of focal subjects is limited, the amount of hours of video recordings on each subject is much higher than anything previously reported in similar settings. While more longitudinal and extensive data collection is certainly necessary, the current data, collected and coded with the same methods recently adopted for chimpanzees and bonobos, allow for a more ecologically valid comparison between the species. We believe this is supporting evidence for the interaction engine hypothesis in that, compared to our closest living relatives, already at a young age humans engage in social interaction more frequently and at a much faster pace than non-human great apes do, interacting with a remarkable number of partners. It remains to be assessed to what degree the limited role played by social experience/partner quantity and strength of social relationships in the unfolding of joint actions in young children is due to more limited social competence, or rather we have overestimated the degree to which adult humans are affected by those same variables in naturally occurring interactions. Ultimately, the data shows that at a very young age humans already manifest a remarkable motivation and skillset for engaging in social interaction that might constitute the hallmark of what distinguishes us from our closest living relatives.

## 5. References

- [1] Schegloff EA. Turn organization: One intersection of grammar and interaction. *Studies in interactional sociolinguistics*. 1996;13:52-133.
- [2] Levinson SC. On the human 'interactional engine'. 2006.
- [3] Levinson SC. Interactional foundations of language: The interaction engine hypothesis. In: *Human language: From genes and brain to behavior*. MIT Press; 2019. p. 189-200.
- [4] Clark HH. *Using language*. Cambridge university press; 1996.
- [5] Bangerter A, Clark HH. Navigating joint projects with dialogue. *Cognitive Science*. 2003;27(2):195-225.
- [6] Sebanz N, Bekkering H, Knoblich G. Joint action: bodies and minds moving together. *Trends in cognitive sciences*. 2006;10(2):70-6.
- [7] Bratman ME. Shared cooperative activity. *The Philosophical Review*. 1992;101(2):327-41.
- [8] Milward SJ, Kita S, Apperly IA. The development of co-representation effects in a joint task: Do children represent a co-actor? *Cognition*. 2014;132(3):269-79.
- [9] Brownell CA. Early developments in joint action. *Review of philosophy and psychology*. 2011;2(2):193-211.
- [10] Gilbert M. Joint commitment. In: Janovic M, Ludwig L, editors. *The Routledge handbook of collective intentionality*. Routledge; 2017. p. 130-9.
- [11] Bangerter this issue?
- [12] Heesen R, Bangerter A, Zuberb"uhler K, Iglesias K, Neumann C, Pajot GE A. Assessing joint commitment as a process in great apes. *IScience*. 2021;24:8.
- [13] Brown P, Levinson SC. *Politeness: Some universals in language usage*. Cambridge University Press; 1987.
- [14] Tomasello M. Joint attention as social cognition. *Joint attention: Its origins and role in development*. ed. by C. Moore and PJ Dunham, 103–30. Hillsdale, NJ: Lawrence Erlbaum; 1995.



- 1  
2  
3 [15] Warneken F, Chen F, Tomasello M. Cooperative activities in young children and  
4 chimpanzees. *Child development*. 2006;77(3):640-63.
- 5 [16] Gräfenhain M, Behne T, Carpenter M, Tomasello M. Young children's understanding of  
6 joint commitments. *Developmental Psychology*. 2009;45(5):1430-43.
- 7 [17] Gräfenhain M, Carpenter M, Tomasello M. Three-year-olds' understanding of the  
8 consequences of joint commitments. *PLoS One*. 2013;8:9.
- 9 [18] Hamann K, Warneken F, Tomasello M. Children's developing commitments to joint goals.  
10 *Child Development*. 2012;83(1):137-45.
- 11 [19] Kachel U, Svetlova M, Tomasello M. Three-year-olds' reactions to a partner's failure to  
12 perform her role in a joint commitment. *Child Development*. 2018;89(5):1691-703.
- 13 [20] Kachel U, Svetlova M, Tomasello M. Three- and 5-year-old children's understanding of how  
14 to dissolve a joint commitment. *Journal of Experimental Child Psychology*. 2019;184:34-47.
- 15 [21] Warneken F, Gräfenhain M, Tomasello M. Collaborative partner or social tool? New  
16 evidence for young children's understanding of joint intentions in collaborative activities.  
17 *Developmental science*. 2012;15(1):54-61.
- 18 [22] Hunnius S, Bekkering H, Cillessen AH. The association between intention understanding  
19 and peer cooperation in toddlers. *International Journal of Developmental Science*.  
20 2009;3(4):368-88.
- 21 [23] Warneken F, Steinwender J, Hamann K, Tomasello M. Young children's planning in a  
22 collaborative problem-solving task. *Cognitive Development*. 2014;31:48-58.
- 23 [24] Rakoczy H, Warneken F, Tomasello M. The sources of normativity: young children's  
24 awareness of the normative structure of games. *Developmental psychology*. 2008;44(3):875.
- 25 [25] Rakoczy H, Schmidt MF. The early ontogeny of social norms. *Child Development  
26 Perspectives*. 2013;7(1):17-21.
- 27 [26] Rossano F, Rakoczy H, Tomasello M. Young children's understanding of violations of  
28 property rights. *Cognition*. 2011;121(2):219-27.
- 29 [27] Keupp S, Behne T, Rakoczy H. Why do children overimitate? Normativity is crucial.  
30 *Journal of Experimental Child Psychology*. 2013;116(2):392-406.
- 31 [28] Brownell CA, Brown E. Peers and play in infants and toddlers. In: *Handbook of social  
32 development*. Springer; 1992. p. 183-200.
- 33 [29] Ross HS, Lollis SP. Communication within infant social games. *Developmental  
34 psychology*. 1987;23(2):241.
- 35 [30] Eckerman CO, Davis CC, Didow SM. Toddlers' emerging ways of achieving social  
36 coordinations with a peer. *Child development*. 1989;440-53.
- 37 [31] Hartup WW, Laursen B, Stewart MI, Eastenson A. Conflict and the friendship relations of  
38 young children. *Child development*. 1988:1590-600.
- 39 [32] Rende RD, Killen M. Social interactional antecedents of conflict in young children. *Early  
40 Childhood Research Quarterly*. 1992;7(4):551-63.
- 41 [33] Hinde RA, Titmus G, Easton D, Tamplin A. Incidence of "friendship" and behavior toward  
42 strong associates versus nonassociates in preschoolers. *Child Development*. 1985:234-45.
- 43 [34] Howes C, Matheson CC. Sequences in the development of competent play with peers:  
44 Social and social pretend play. *Developmental Psychology*. 1992;28(5):961.
- 45 [35] Bakeman R, Adamson LB. Coordinating attention to people and objects in mother-infant  
46 and peer-infant interaction. *Child Development*. 1984:1278-89.
- 47 [36] Brownell CA. Early developments in joint action. *Review of philosophy  
48 and psychology*. 2011;2(2):193-211.
- 49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 [37] Hay DF, Caplan M, Nash A. The beginnings of peer relations. Handbook of peer  
4 interactions, relationships, and groups. 2018;200-21  
5 [38] Altmann J. Observational study of behavior: sampling methods. Behavior. 1974;49(3-  
6 4):227-66.[39] Rubin KH, Watson KS, Jambor TW. Free-play behaviors in preschool and  
7 kindergarten children. Child development. 1978;534-6.  
8 [40] Lillard A, Pinkham AM, Smith E. Pretend play and cognitive development. In: Goswami  
9 BU, editor. The Wiley-Blackwell handbook of childhood cognitive development. 285–311-:  
10 Wiley; 2011.  
11 [41] Shantz CU, Shantz DW. Conflict between children: Social-cognitive and sociometric  
12 correlates. New directions for child development. 1985.  
13 [42] Dawe HC. An analysis of two hundred quarrels of preschool children. Child Development.  
14 1934;5(2):139-57.  
15 [43] Hay DF, Ross HS. The social nature of early conflict. Child development. 1982:105-13.  
16 [44] Jersild AT, Markey FV. Conflicts between preschool children. Child Development. 1935  
17 [45] Dunn J, Herrera C. Conflict resolution with friends, siblings, and mothers: A developmental  
18 perspective. Aggressive Behavior: Official Journal of the International Society for Research on  
19 Aggression. 1997;23(5):343-57.  
20 [46] Killen M, Turiel E. Conflict resolution in preschool social interactions. Early education and  
21 development. 1991;2(3):240-55.  
22 [47] Shantz CU. Conflicts between children. Child development. 1987:283-305.  
23 [48] Corsaro WA, Eder D. Children's peer cultures. Annual review of sociology.  
24 1990;16(1):197-220.  
25 [49] O'Brien M, Roy C, Jacobs A, Macaluso M, Peyton V. Conflict in the dyadic play of 3-year-  
26 old children. Early Education and Development. 1999;10(3):289-313.  
27 [50] Vespo JE, Pedersen J, Hay DF. Young children's conflicts with peers and siblings: Gender  
28 effects. Child Study Journal. 1995.  
29 [51] Heesen R, Bangerter A, Zuberbuhler K, Rossano F, Iglesias K, Guery JP, et al. Bonobos  
30 engage in joint commitment. Science Advances. 2020;6:51.  
31 [52] ELAN (Version 6.3) [Computer software]. (2022). Nijmegen: Max Planck Institute for  
32 Psycholinguistics, The Language Archive. Retrieved from <https://archive.mpi.nl/tla/elan>  
33 [53] Byrne RW, Cartmill E, Genty E, Graham KE, Hobaiter C, Tanner J. Great ape gestures:  
34 intentional communication with a rich set of innate signals. Animal cognition. 2017;20(4):755-  
35 69.  
36 [54] Hill KR, Wood BM, Baggio J, Hurtado AM, Boyd RT. Hunter-gatherer inter-band  
37 interaction rates: Implications for cumulative culture. PloS one. 2014 Jul 21;9(7):e102806.  
38 [55] Hay DF. Social conflict in early childhood. Annals of Child Development. 1984;1:1-44.  
39 [56] Houseman JA. An ecological study of interpersonal conflicts among preschool children.  
40 Wayne State University; 1972.  
41 [57] Kaufhold SP, Rossano F. Proximate mechanisms and relational history: the interdependence  
42 of food transfers in socially housed orang-utans (*Pongo abelii*). Animal Behaviour.  
43 2020;167:243-53.  
44 [58] Rossano F, Liebal K. Requests' and 'offers' in orangutans and human infants. Requesting in  
45 social interaction. 2014 Dec 15:333-62.  
46 [59] Schegloff EA, Sacks H. Opening up closings. Semiotics. 1973;8:289-327.  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60