

Social Context Matters for Turn-Taking Dynamics: A Comparative Study of Autistic and Typically Developing Children

Christopher Cox^{*1,2}, Riccardo Fusaroli^{*1,2,3}, Yngwie A. Nielsen^{1,2}, Sunghye Cho^{3, 4}, Roberta Rocca^{1,2}, Arndis Simonsen^{2,5}, Azia Knox⁶, Meg Lyons⁷, Mark Liberman^{3,4}, Christopher Cieri³, Sarah Schillinger⁷, Amanda L Lee⁷, Aili Hauptmann⁷, Kimberly Tena⁸, Christopher Chatham⁹, Judith S. Miller^{7,10,11}, Juhi Pandey^{7,10,11}, Alison S. Russell¹², Robert T. Schultz^{7,10,11}, Julia Parish-Morris^{7,10,11}

¹Department of Linguistics, Cognitive Science and Semiotics, Aarhus University

²Interacting Minds Center, Aarhus University

³Linguistic Data Consortium, University of Pennsylvania

⁴Department of Linguistics, University of Pennsylvania

⁵Psychosis Research Unit, Psychiatry, Aarhus University Hospital

⁶Department of Linguistics and Cognitive Science, University of Delaware

⁷Center for Autism Research, Children's Hospital of Philadelphia

⁸Carolina Autism and Neurodevelopment Research Center, University of South Carolina

⁹Hoffmann-La Roche Ltd

¹⁰Department of Psychiatry, Perelman School of Medicine at the University of Pennsylvania

¹¹Department of Child and Adolescent Psychiatry and Behavioral Sciences, Children's Hospital of Philadelphia

¹²Department of Psychology and Neuroscience, University of North Carolina Chapel Hill

2025

Authors marked with * share joint first-authorship. Correspondence to: Christopher Cox
(chris.mm.cox@gmail.com)

Abstract

Engaging in fluent conversation is a surprisingly complex task that requires interlocutors to promptly respond to each other in a way that is appropriate to the social context. In this study, we disentangled different dimensions of turn-taking by investigating how the dynamics of child-adult interactions changed according to the activity (task-oriented versus freer conversation) and the familiarity of the interlocutor (familiar versus unfamiliar). Twenty-eight autistic children (16 male; $M_{age} = 10.8$ years) and 20 age-matched typically developing children (8 male; $M_{age} = 9.6$ years) participated in seven task-orientated face-to-face conversations with their caregivers (336 total conversations) and seven more telephone conversations alternately with their caregivers (144 total conversations, 60 with the typical development group) and an experimenter (191 total conversations, 112 with the autism group). By modelling inter-turn response latencies in multi-level Bayesian location-scale models, we found that inter-turn response latencies were consistent across repeated measures within social contexts, but exhibited substantial differences across social contexts. Autistic children exhibited more overlaps, produced faster response latencies and shorter pauses than typically developing children – and these group differences were stronger when conversing with the unfamiliar experimenter. Unfamiliarity also made the relation between individual differences and latencies evident: only in conversations with the experimenter were higher socio-cognitive skills and lower social awareness associated with faster responses. Information flow and shared tempo were also influenced by familiarity: children adapted their response latencies to the predictability and tempo of their interlocutor's turn, but only when interacting with their caregivers and not the experimenter. These results highlight the need to construe turn-taking as a multicomponential construct that is shaped by individual differences, interpersonal dynamics, and the affordances of the context.

1. INTRODUCTION

Humans navigate their social world through conversations. Despite their apparent ease and naturalness, fluent conversations are complex feats of coordination involving rapid responses that operate at the absolute limits of human reaction times. The typical gap between conversational turns lies at approximately 59–200 ms (Dingemanse & Liesenfeld, 2022; Levinson, 2016; Stivers et al., 2009), which is too fast to rely on reactive planning and production after the end of the previous turn (Magyari et al., 2014; Magyari & De Ruiter, 2012). Interlocutors must therefore proactively prepare their responses, predict the current turn end and engage their articulators ahead of time (Dale et al., 2013; Magyari & De Ruiter, 2012; Magyari et al., 2014). While adults draw upon their linguistic, motor and interpersonal skills as well as world knowledge to plan and predict with such efficiency (Ford et al., 1996; Magyari & De Ruiter, 2012; Magyari et al., 2014), it remains an open question how children manage this complex act of coordination. In fact, response latency has been taken as a simple (one-to-one) proxy of social competence without regard for the contextual and intraindividual processes that underlie turn-taking. In this paper, we investigate how child-adult turn-taking dynamics are influenced by the social context, the moment-to-moment dynamics of the conversation, and the individual differences in linguistic, socio-cognitive and motor skills of both typically and atypically developing children.

1.1. Turn-taking as Social Competence

Turn-taking emerges early in ontogeny (Levinson, 2016), is exhibited by our nearest phylogenetic ancestors and across the animal kingdom (Ravignani & de Reus, 2019; Levinson, 2006, 2016; Verga et al., 2023; Okobi Jr et al., 2019; Cox, Templeton, & Fusaroli, 2023), and displays similar structures across languages (Stivers et al., 2009; Dingemanse & Liesenfeld, 2022). To explain this ubiquity, turn-taking has been argued to arise from a "package" of fundamental capacities enabling human social behaviour (Levinson, 2006, 2016). Accordingly, turn-taking has important social implications; shorter inter-turn latencies facilitate efficient information exchange (R. Fusaroli, Rączaszek-Leonardi, & Tylén, 2014; Templeton et al., 2022, 2023) and foster a sense of connection and shared motivation (Dideriksen, Christiansen, Tylén, et al., 2023; R. Fusaroli et al., 2019; R. Fusaroli & Tylén, 2016; Konvalinka & Roepstorff, 2012; Tylén et al., 2013). Adults in verbal interactions with shorter inter-turn response latencies, for example, report enjoying their conversations more (Templeton et al., 2022).

In light of these findings, there is a broad tendency to equate short turn-taking latencies with skillful turn-taking, and in turn, to equate individual variation in turn-taking latency to the expression of social competence (R. Fusaroli et al., 2025; Cho et al., 2023). However, the shortcomings of this perspective are increasingly becoming apparent, especially within developmental and clinical work. Several empirical studies have documented slower turn-taking latencies in children with autism (W. Fay & Schuler, 1980; Heeman et al., 2010; Ochi et al., 2019; Warlaumont et al., 2014), leading to the mainstream conception that autistic children are slower due to a social competence deficit (see Nguyen et al. (2022) for a systematic review). However, recent studies have either failed to replicate this finding (Cho et al., 2023; Wehrle et al., 2023) or obtained the opposite pattern, namely faster turn-taking in autistic children (R. Fusaroli et al., 2025). Part of the explanation lies in the observation that children with autism exhibit both larger overlaps (i.e., instances of starting to speak before the previous speaker is done) and longer-tailed distributions (i.e., instances of very long gaps), suggesting that autistic children may initiate their turns faster or slower than typically developing children, depending on the context (R. Fusaroli et al., 2025). These findings dovetail with qualitative investigations showing that autistic children generally do not display difficulties in conversational turn-taking with caregivers (Ochs et al., 2004; Bottema-Beutel et al., 2018), possibly due to caregivers scaffolding turn-taking by producing utterances that invite contingent responses (Bottema-Beutel et al. (2018); Bottema-Beutel et al. (2022); see also Dunn & Shatz (1989); Ervin-Tripp (1982) for evidence in typically developing children). Further, developmental trajectories can be very different from child to child, and this heterogeneity is even more prominent in autism (R. Fusaroli et al., 2019). Contrasting autistic and typically developing children as groups risks neglecting more proximal causes of differences in turn-taking behaviours, such as crucial variation in linguistic, socio-cognitive, and motor abilities, and how they play out in different contexts. These findings suggest the need to look beyond turn-taking as a monolithic construct, focusing instead on how turn-taking emerges from the interplay of individual differences, interpersonal dynamics and contextual demands.

1.2. Unfolding the Components of Turn-Taking

No two conversations are the same. A relaxed chat and an oral examination posit wildly different contextual demands on the conversation, including the complexity of the linguistic content to plan and process as well as what is considered an appropriate turn-taking latency. Likewise, no two individuals are the same, and turn-taking dynamics likely depend on several linguistic, socio-cognitive and motor skills. In the following, we construe the temporal dynamics of turn-taking as a window into this complex network of interacting factors and argue that a nuanced perspective on verbal interactions requires consideration of the following components: the socio-cognitive, linguistic and motor skills of the speaker, the properties of the individual utterances, and the unfolding interpersonal processes of the speakers. How these individual components play out in turn-taking depends on contextual demands, including the purpose of the conversation and the relation between interlocutors.

1.2.1 Purpose of the Conversation

Verbal interactions do not happen in a vacuum, but emerge as an intrinsic part of joint activities with distinct affordances and constraints (R. Fusaroli, Rączaszek-Leonardi, & Tylén, 2014; R. Fusaroli & Tylén, 2012; Cola et al., 2022; Dale, 2015; R. Fusaroli, Gangopadhyay, & Tylén, 2014; Parish-Morris et al., 2019; Olsen & Tylén, 2023). These affordances and constraints may directly affect turn-taking dynamics. For instance, a joint task that requires children and parents to tell each other where to place stickers on a blank picture demands a focus on precise information transfer. More importantly, it requires interlocutors to allow each other sufficient time to place the stickers before responding. In contrast, the situational demands of a spontaneous conversation about holidays demands less precision, presumably resulting in fewer long pauses (see Dideriksen, Christiansen, Tylén, et al. (2023); Dideriksen, Christiansen, Dingemanse, et al. (2023); Dideriksen et al. (2019); R. Fusaroli, Tylén, et al. (2017) for examples of asymmetric access to knowledge, or asymmetric roles).

1.2.2 Relation to the Interlocutor

Equally important is the person we interact with in joint activities. Different interlocutors share different levels of common ground (i.e., assumed common knowledge; (H. H. Clark, 1996; Branigan et al., 2016; De Marchena & Eigsti, 2016)), which provides the conditions for transferring information. For instance, chatting with a long-known friend is very different from talking with a stranger at a party, and discussing which toy to bring at a show-and-tell at school is very different if doing so with a peer compared to a teacher. In child-adult interactions, the familiarity of the interlocutor changes how both autistic and typically developing children interact (Doherty-Sneddon et al., 2013; Dawson et al., 2002; Forgeot d'Arc et al., 2020). Neurotypical adults interacting with friends likewise take longer to respond compared to when interacting with strangers, and their occasional long pauses are perceived more positively (Templeton et al., 2022, 2023).

Turn-taking patterns in child-adult conversations, then, represent an interaction between the structural demands of the task and the degree of familiarity with an interlocutor. The study design of the current paper allowed us to evaluate how different affordances emerging from the task and the familiarity of the interlocutor produced differences between the autism

and typical development group in their turn-taking behaviours (see Research Question 1 & 2 below). Children and adults were recorded in each of these social contexts in multiple weekly sessions to disentangle heterogeneity due to session-by-session variation from systematic contextual variation.

1.2.3 Individual Differences

The challenges posed by various contextual demands and interlocutors require a complex network of skills to overcome. For instance, promptly responding requires mature linguistic skills; that is, the ability to integrate semantic, syntactic and pragmatic information to anticipate an imminent termination of a turn (Magyari & De Ruiter, 2012). According to the *Interaction Engine Hypothesis*, human socio-cognitive skills lie at the core of turn-taking (Levinson, 2016), suggesting a role for individual differences in social skills. Thus, smooth turn-taking also requires the speaker to infer others' intentions, anticipating what they might do next, and plan behaviours that are predictable for the interlocutor (Heesen & Fröhlich, 2022; Levinson, 2006; Waade et al., 2023). In addition, motor development may be important as it affects the nature of children's interactions with the world around them, making them better at promptly anticipating, orienting towards, and reacting to information sources (Iverson, 2010). These skills have crucial downstream effects on their own communicative abilities (West & Iverson, 2021), as well as on the contingent responses from caregivers (Karasik et al., 2014; Tamis-LeMonda & Adolph, 2013). Besides directly affecting the children's behaviours, individual differences in skills changes the input that children receive from caregivers; several studies of adults have shown that they adapt to their addressees, including non-native speakers (Uther et al., 2007; Piazza et al., 2022), autistic and non-autistic children (Cox, Bergmann, et al., 2023; Hilton et al., 2022; Neimy et al., 2017; Warlaumont et al., 2014; Leezenbaum et al., 2014), infants with cochlear implants (Dilley et al., 2020; Kondaurova et al., 2013), or pets (Jardat et al., 2022; Panneton et al., 2023).

Crucially, different contextual and interlocutor demands might engage these individual skills to a higher or lower degree. For instance, the behaviour of familiar interlocutors might be easier for children to anticipate and respond to, while the ability to promptly take turns with an unfamiliar interlocutor might be more dependent on socio-cognitive skills (Bigelow, 1999; Jaffe et al., 2001). In other words, assessing how individual differences relate to response latencies as we vary the characteristics of the social context is essential to improving our understanding of their role in turn-taking. To assess these ideas, our study incorporated rich measures of individual differences to examine how socio-cognitive, linguistic and motor differences among children might change the turn-taking dynamics of interactions across social contexts with different task demands (see Research Question 3 below).

1.2.4 Properties of the utterance

Conversations are dynamic, and conversational demands vary on a turn-by-turn basis in terms of the novelty of information, the effort required to parse what has been said, and the effort required to plan the next utterance. In adults, inter-turn response latencies depend on the predictability of the content discussed (Bögels et al., 2015), and preschool children produce shorter inter-turn response latencies with more predictable ends to questions (Lindsay et

al., 2019; Casillas & Frank, 2017). Intuitively, children having to respond to a complex, open-ended question (e.g., "What are your thoughts on the story we just read?") will likely produce longer response times compared to a predictable, closed-ended question (e.g., "Do you have an apple in your hand?").

To assess this component, our study quantified two distinct aspects of predictability. First, we measured the processing effort required for an incoming utterance by calculating how predictable it was from previous turns (to investigate whether more predictable utterances require less processing effort, potentially enabling faster responses). Second, we measured the planning effort for an outgoing utterance by calculating how predictable the speaker's upcoming utterance was compared to previous turns. For this planning aspect, we investigated whether more predictable utterances might actually require more or less coordination with the ongoing conversation flow. On the one hand, highly predictable responses might demand closer attention to the conversational context and more careful coordination, possibly resulting in longer planning time. On the other hand, highly predictable responses might build on previous utterances and therefore require less effort and cognitive load. We then assessed whether these complementary measures of predictability had reliable associations with inter-turn response latencies (see Research Question 4 below).

1.2.5 Interpersonal dynamics

Beyond pure exchanges of information, conversations are also a social dance in which interlocutors might come to share a common tempo of well-timed exchanges (Wilson & Wilson, 2005; R. Fusaroli, Rączaszek-Leonardi, & Tylén, 2014). For example, in a conversation between a child and caregiver during a routine activity (e.g., preparing a meal), the child might naturally match the caregiver's rhythm, responding with pauses that mirror the caregiver's slower pace. However, during a more structured activity, like a reading lesson, the caregiver might adopt a faster tempo, prompting the child to respond more quickly. This shift in tempo is not just a matter of individual response speed but reflects an ongoing process of adaptation, where both participants adjust their timing in response to the conversational and social context, as well as to each other (Pouw & Holler, 2022; Zamm et al., 2015). There are general findings of strong correlations between child and caregiver response latencies (Nguyen et al., 2022); however, these findings are based only on summary estimates (the average response latency across the whole conversation) and cannot say much about the dynamics of co-adaptation on a turn-by-turn basis, nor how context demands or individual differences influence such co-adaptation (R. Fusaroli et al., 2025). Further, vocal interactions likely involve dynamic ebb-and-flow waves of mutual convergence and scaffolding that takes place during local transactions (Ritwika et al., 2020; Warlaumont et al., 2014; Pouw & Holler, 2022), but little is known about how this unfolds in child-adult interactions. Accordingly, our study included an analysis of these temporal dependencies: how child and caregiver adjust to each other's - and their own - previous latencies on a turn-by-turn basis across contexts (see Research Question 5 below).

1.3. The present research

In the present research, we unpack different components of turn-taking and investigate how they unfold across diverse social contexts, in autism and typical development; and unravel contradictory results pertaining to turn-taking in autistic children. Specifically, we focus on the following five research questions (RQs):

- **RQ1:** Are response latencies an individual trait, or do differences between sessions and social contexts (conversational task and familiarity of the interlocutor) matter?
- **RQ2:** Does autism on average involve slower response latencies than typical development - once explicitly accounting for long pauses and overlaps? How is the finding shaped by context?
- **RQ3:** Are differences in turn-taking behaviour grounded in individual differences (social cognition, social awareness and social motivation, language skills and motor skills)? How are the findings shaped by context?
- **RQ4:** Is turn-taking behaviour across social contexts grounded in the turn-by-turn information flow (predictability of previous and current utterance)?
- **RQ5:** Is turn-taking behaviour across social contexts grounded in the turn-by-turn temporal adjustments in the interaction?

We study these RQs using a corpus of child-adult conversations with both neurotypical and autistic children, several contexts (i.e., free and task-orientated conversations as well as familiar and unfamiliar interlocutors), measures of individual differences (i.e., social, cognitive, motor skills), information flow (i.e., utterance predictability) and turn-by-turn temporal adjustments (i.e., self and interpersonal adjustment).

2. METHODS

2.1. Participants

The study involved 28 autistic (10 female, 2 non-binary) and 20 typically developing children (12 female) children, matched at the group level on age, IQ and self-reported race (see **Table 1**). All autistic participants had previously received a clinical diagnosis of Autism Spectrum Disorder (ASD) from licensed psychologists or physicians external to the research centre. When available, previous documentation of diagnostic reports, previous evaluations, and Individualized Education Program (IEP) documentation was requested. While some diagnostic documentation may have been based on earlier frameworks (i.e., DSM-IV), the confirmatory evaluation procedures aligned with DSM-5 criteria. Such diagnostic confirmation at the research centre included the Autism Spectrum Disorder Diagnostic Interview and Activities-Lifespan (ASD-DIAL), the Childhood Autism Rating Scale, 2nd Edition (CARS-2), standardized cognitive and language assessments, and parent-reported questionnaires (e.g., SCQ, SRS-2) to confirm ASD diagnosis or lack thereof. Participants were recruited via an electronic health record system at a large hospital-based academic medical center in the U.S.,

flyers and inquiry forms, and internal databases. All participants underwent the Telephone Screening Interview (TSI) developed by the researchers at the Center for Autism Research to ascertain adherence to inclusion and exclusion criteria (such as autism diagnostic history, family history, or upcoming changes to intervention/medication). The inclusion criteria were as follows: all participants had to (i) be between 6 and 18 years of age, (ii) have English as their primary language, (iii) be verbally fluent consistent with their chronological age, (iv) have an IQ score above 75, and (v) have a current Social and Communication Questionnaire (SCQ) score above 10 (for autistic children) or below 11 (for neurotypical children), where higher scores represent a higher degree of social communication impairment (Berument et al., 1999; Chesnut et al., 2017). Verbal fluency was assessed through standardized language evaluations and clinical judgment to ensure all participants could engage meaningfully in conversational tasks. The exclusion criteria were as follows: all participants had (i) no known genetic or neurological condition that impacts neurodevelopment, language, or vocal production, (ii) no history of expressive language deficits or severe neurological injury likely to affect expressive language and communication behavior, (iii) no extreme birth prematurity (<32 weeks), (iv) no diagnosed hearing impairment or cochlear implant, (v) no plan to begin or change medication or intervention during study duration, and (vi) no first-degree family members with autism (for the neurotypical children only).

2.2. Procedure

2.2.1 Data Collection

Participant data were collected over the course of several telephone sessions. After an information call and having provided consent, the participants and their families underwent the screening questionnaire (TSI) and the Social and Communication Questionnaire (Berument et al., 1999; Chesnut et al., 2017). Subsequent assessment relied on the Wechsler Abbreviated Scale of Intelligence, 2nd Edition (WASI-II) (Wechsler, 2008; McCrimmon & Smith, 2013), the Clinical Evaluation of Language Fundamentals, 5th Edition (CELF-5) (Wiig et al., 2013; Denman et al., 2017), the Vineland Adaptive Behavior Scales, 3rd Edition (Vineland-3) (Sparrow et al., 2016; Farmer et al., 2020), and the Social Responsiveness Scale (SRS-2), 2nd edition (Constantino, 2012; Bruni, 2014). While the conversational data were collected via telephone, all standardized assessments (WASI-II, CELF-5, Vineland-3, SRS-2) were administered during in-person research visits by trained clinicians following standard administration protocols.

Table 1: Descriptive statistics of the sample.

Variable	Autism Group	Typical Development
Gender	10 female, 16 male, 2 non-binary	12 female, 8 male
Age (years)	10.82 (3.14)	9.6 (3.02)
Age Range (years)	6-17	6-15
IQ	112.88 (13.26)	118.16 (13.15)
SCQ	16.68 (16.68)	1.20 (1.20)
SRS (Raw)	77.57 (25.99)	12.75 (9.13)
SRS (T-score)	67.77 (9.92)	42.50 (3.44)
Language (CELF-5)	107.67 (13.15)	109.53 (11.40)
Social Cognition (SRS-2)	13.71 (5.46)	1.8 (2.21)
Social Motivation (SRS-2)	12.36 (6.11)	3 (2.13)
Social Awareness (SRS-2)	10.39 (3.51)	2.5 (2.04)
Motor (VAB-3)	82.54 (3.82)	85.25 (2.07)
Number of Children	28	20
Number of Sessions	392	279
Number of Turns	21733	14628
Average Turns Per Session	27.7	26.3
Median Turn Length (seconds)	2.87 (8.22)	2.69 (9.93)
Mean Turn Length (seconds)	5.69 (8.22)	5.97 (9.93)

Note. Numbers in parentheses denote standard deviations for each of the variables. The five last rows display details about the turns for each of the groups under investigation. Median and Mean Turn Length refers to the average difference in seconds between the startpoint and endpoint of children's turns across conditions. IQ = Intelligence Quotient, as measured with the Wechsler Abbreviated Scales of Intelligence – 2nd Edition. SCQ = Social and Communication Questionnaire. SRS-2 = Social Responsiveness Scales – 2nd Edition. CELF-5 = Clinical Evaluation of Language Fundamentals - 5th edition. VAB-3 = Vineland Adaptive Behavior - 3rd edition.

After evaluation according to the above assessment tools, participants underwent phone-based recording sessions approximately once per week for seven weeks. The participants participated in two speech recording tasks, namely a matching game (only performed with the parent) and a prompted conversation (alternately with the parent and with an unfamiliar research assistant). These two tasks were adapted from established paradigms in conversational research (Dideriksen, Christiansen, Tylén, et al., 2023; N. Fay et al., 2018) to be engaging and repeatable for our age range (6-17 years). In the matching game, the child and parent each received a sticker sheet and a blank picture and were instructed to take turns in telling the other person where to place the stickers. The aim of the game was to obtain identical pictures despite not being able to see each other's sticker sheet. In the prompted conversations, children were given guided prompts (e.g., birthday party planning, hobbies, food preferences) and interacted alternately with their caregivers in even weeks and with an initially unfamiliar experimenter in odd weeks. Participants were asked to chat for approximately 5 minutes, though actual durations varied naturally (see **Table 2**). The matching game with parents typically lasted longest, while parent conversations were slightly briefer than those with

experimenters.

All recording sessions occurred in participants' homes with caregivers present to assist children in conversations with the experimenter when needed. During the first session with the child, the research assistant remained on the line to facilitate tasks with the parent and child, while in following sessions children and parents completed the tasks on their own. We aimed to maintain consistent research assistant assignments for each participant across all sessions, but this was not always feasible due to scheduling constraints and research assistant availability (see **Figure S22** for an overview of research assistant assignments for each participant and **Figure S23** for a control analysis). Phone calls were scheduled based on family availability, and we tried to maintain consistency in session times where possible. The sequence of tasks was not counterbalanced across visits or participants; all sessions followed the same standardized protocol where participants first completed the matching game followed by the prompted conversation. This consistent ordering was maintained across all seven weekly visits and all participants to ensure procedural consistency. All technical issues were systematically documented, with detailed session-by-session breakdowns provided in **Table S15**. Regarding the sequence of conversations, participants spoke with unfamiliar partners (research assistants) during odd-numbered weeks and caregivers during even-numbered weeks, completing both activities (matching game and conversation) within the same phone call when speaking with caregivers. This experimental setup implies that we have the following number of sessions for each child: 7 sessions for the matching game with parents, 4 sessions of conversations with the experimenter and 3 sessions of conversations with the parent.

Table 2: *Average Task Duration (in minutes)*

Task	Autism Group	Typical Development
Matching With Parent	18.7 (7.44)	17.3 (6.45)
Convo With Parent	3.52 (1.62)	3.50 (1.48)
Convo With Experimenter	5.25 (3.14)	4.89 (2.02)

Note: Values represent mean durations in minutes with standard deviations in parentheses.

2.2.2 Recording and Transcribing Turns

To record participants' speech we relied on an automated phone bank collection system developed by the University of Pennsylvania Linguistic Data Consortium and adapted for use with autistic participants and across a wide variety of different devices. The speech recordings were transcribed by trained annotators using a web-based transcription tool with a built-in speech activity detector function. Data transcription involved annotators assessing and correcting the timestamps for automatically identified speech segment boundaries. Analyses of inter-transcriber reliability of speech segment boundaries demonstrated excellent agreement ($\kappa = 0.88$ [0.74, 0.97] for conversations and $\kappa = 0.88$ [0.76, 0.94] for the matching game, with percentage agreement across all recordings was 94.7%, see Supplementary Materials). Conversational turns were conceptualized as an idealized ABABAB structure where each speaker takes sequential turns. In practice, overlapping speech sometimes occurred, which we handled through specific segmentation procedures to preserve the sequential analysis frame-

work. When both interlocutors produced speech simultaneously, we combined backchannels with the subsequent turn by the same interlocutor, calculating latency based on the timing of substantial content following the backchannel rather than the backchannel onset itself. This approach maintained the ABABAB structure while accounting for natural conversation dynamics. Utterances with negative latencies thus represent cases where a full utterance (not a backchannel) began before the previous speaker had completed their turn.

2.3. Operationalizing Turn-Taking

Inter-turn response latencies were calculated by subtracting the onset of an utterance from the offset of the previous utterance. This yielded both negative (overlaps) and positive response latencies (gaps), including a non-trivial number of long pauses. In previous research, overlaps have often been excluded as failures of communication, or have been considered to be qualitatively different from gaps (Nguyen et al., 2022; R. Fusaroli et al., 2025; Ochi et al., 2019). Long pauses have likewise either been included but ignored (i.e., using likelihood functions without long right tails), or been excluded as outliers at various thresholds changing from study to study (Gratier et al., 2015; E. V. Clark & Lindsey, 2015). These different conceptions of turn-taking can lead to different findings even when using the same dataset. For example, two studies explicitly including overlaps found that response latencies for the autism group were equivalent or faster than those of the typical development group (R. Fusaroli et al., 2025; Wehrle et al., 2023), contrary to studies excluding overlaps (Ochi et al., 2019; Nguyen et al., 2022).

In this study, we chose to include overlaps and long pauses and to explicitly account for them in the statistical model. We made this choice because overlaps constitute a substantial portion of conversations: between 25% and 50% of conversational utterances are estimated to start before the previous interlocutor is done with speaking (Dingemanse & Liesenfeld, 2022), and this is argued to be even more frequent in conversations involving children (Ervin-Tripp, 1979; Garvey & Berninger, 1981). Moreover, there is a growing recognition that overlaps may serve several different functions, similar to how long gaps may signify both thoughtful engagement of the interlocutor or a desire to avoid a topic (Templeton et al., 2023). Specifically, overlaps can indicate disengaging from the conversation and interruptions, but also shared attention and mutual understanding (as in backchannels), completion of the interlocutor's utterance, or they may signal disagreement and the need for potential repair (Dideriksen, Christiansen, Tylén, et al., 2023). Overlaps thereby play a crucial role in building and maintaining rapport (Bryant et al., 2016; Bryant, 2020; Templeton et al., 2022; Cummins, 2019). Together with long pauses, overlaps form a natural part of smooth turn-taking, and, we argue, should be integrated into models of turn-taking instead of being discarded. Beyond the conceptual reasons for including overlaps, removing overlaps and long pauses also truncates an otherwise continuous distribution, which violates traditional distributional assumptions (R. Fusaroli et al., 2025). To nonetheless allow for comparisons with prior studies (for a review see Nguyen et al. (2022) and highlight the impact of our modeling choices, we conducted additional analyses following the prior work (i.e., excluding overlaps, no explicit modelling of long pauses). We report these in the Supplementary Materials.

2.4. Statistical Models

The distributional properties of response latencies in turn-taking produce unique modelling challenges (see **Figure 1**). Turn-taking distributions involve negative latencies from overlapping speech, a large bulk of observations just above zero, and positive skew from occasional long pauses between turns (R. Fusaroli et al., 2025; Heldner & Edlund, 2010). To accommodate these distributional properties in our models, we applied Bayesian multilevel models with ex-Gaussian likelihoods for all of our research questions. This form of likelihood combines a Gaussian distribution with an exponential distribution, allowing for the flexibility to represent the central tendency together with the positive skew often observed in empirical turn-taking data. Specifically, in our Bayesian framework, the ex-Gaussian likelihood was parameterized by three parameters: the mean (μ) and variance (σ) of the Gaussian component, as well as the rate (β) of the exponential tail. Control analyses using simpler distributional assumptions (see Supplementary Materials) demonstrated that failing to account for the exponential tail component would lead to biased estimates of central tendency, particularly in contexts with more frequent long pauses such as the matching game. The ex-Gaussian explicitly models these distributional characteristics, allowing us to separately quantify differences in typical response timing from differences in the propensity for long pauses, thereby providing a more complete picture of turn-taking dynamics across groups and contexts.

To investigate whether response latencies are an individual trait (RQ1) and whether autism on average involves slower response latencies than typical development (RQ2), we conditioned each of the parameters - the Gaussian mean (μ) and variance (σ) as well as the exponential rate (β) - on the diagnostic group (neurotypical vs. autistic children) and conversational context (matching game vs. conversations with parents vs. conversations with experimenters). To assess group differences across the tasks and familiarity of the interlocutor, we allowed the estimates to vary within each combination of these factors (similar to a full interaction model). To account for the hierarchical structure of our data, we implemented a multilevel Bayesian model with random effects at both participant and visit levels. At the participant level, random slopes were included for each Task-Familiarity combination. At the visit level, random slopes captured how the effects of different Task-Familiarity combinations varied across the seven weekly sessions. While not all combinations occurred in every visit (e.g., conversations with unfamiliar partners only occurred in odd-numbered weeks), this structure accounted for visit-specific effects that might influence participants' turn-taking behaviors over time. Variance was partially pooled by child within their respective diagnostic group to accommodate the potential for between-group differences in heterogeneity in mean and variance. To investigate test-retest reliability of our measures of response latency, we calculated correlations between individual participants' mean latencies across repeated visits within each social context (all seven visits for matching game; visits 2, 4, 6 for parent conversations; visits 1, 3, 5, 7 for experimenter conversations) with bootstrapped confidence intervals ($N = 1000$). We also assessed cross-context consistency by correlating each child's mean latency estimates between the three social contexts to examine whether individual differences in response timing represent stable characteristics that persist across different social situations. More information about the specification of our models, including priors, model checks and control analyses can be found in Supplementary Materials.

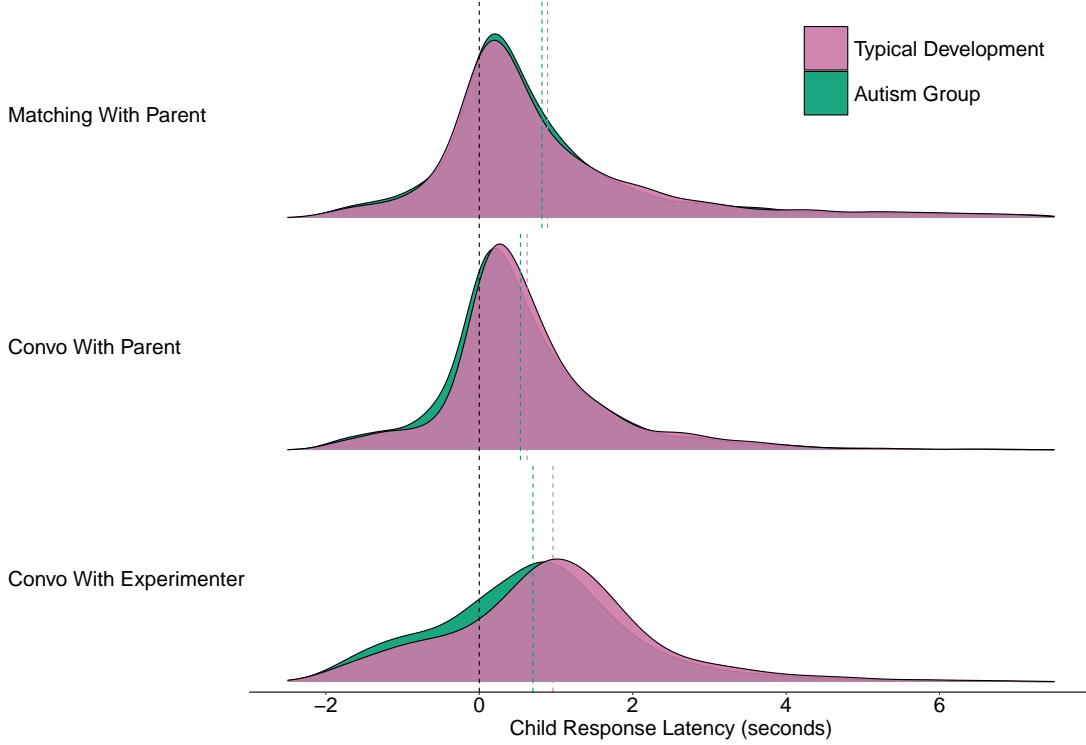


Figure 1: Plots to demonstrate the distributional properties of child response latencies across different combinations of contexts and familiarity. The black dashed line marks a response latency of zero, the purple dashed line the mean in the typical development group, and the green dashed line the mean of the autism group. As can be gleaned from the plots, response latencies involve both negative response latencies (overlapping speech), a large centre of mass just above zero, as well as positive skew (long pauses between turns)

To investigate the role of individual differences on turn timing (RQ3), we added scaled scores of SRS-2 (Social Cognition, Social Motivation, Social Awareness), CELF-5 (Language), and VAB-3 (Motor) to the first model (see **Table 1**). The Vineland Adaptive Behavior Scales, Third Edition (VAB-3) is standardized for individuals from birth through 90 years of age, with age-appropriate norms that allow for meaningful comparisons across our participants despite age differences. We should note that while we used general motor skills from the Vineland Adaptive Behavior Scales (VAB-3) to provide a broad assessment of motor abilities, a limitation of this approach is the absence of specific oral-motor or speech-motor measures, which might be more directly relevant to turn-taking mechanisms given recent evidence linking speech-motor networks to timing differences in autism (O'Brien et al., 2023). Rather than treating social skills as a unitary construct, moreover, we distinguished between three key dimensions from the SRS-2 inventory: social cognition, social awareness, and social motivation (Constantino, 2012; Bruni, 2014). While we acknowledge that the scaled scores from three subscales of the SRS-2 were originally designed for clinical reference rather than as independent research variables, they provide useful constructs for examining different aspects of social functioning. These distinct but related dimensions of social functioning may contribute differently to turn-taking dynamics in conversation. Social awareness involves the ability to perceive social cues (e.g., noticing changes in facial expressions or shifts in

tone of voice), while social cognition refers to the ability to interpret social cues (e.g., understanding that a person leaning forward indicates interest or that a brief silence might signal confusion). Social motivation captures the drive to engage in social interaction and form interpersonal bonds. These distinct but related dimensions of social functioning may contribute differently to turn-taking dynamics in conversation. For example, social awareness might enhance detection of turn-yielding cues, social cognition might facilitate anticipation of turn completions, and social motivation might influence the effort devoted to maintaining conversational timing. Their relation to response latencies was allowed to vary by diagnostic group, conversational context, and familiarity (equivalent to a full interaction model), enabling us to assess how these relationships differed across contexts. We chose to use SRS scores of socio-cognitive skills instead of Vineland Socialization scores since individual abilities were measured only once during data collection. Vineland scores are argued to be more sensitive to small variations between weekly sessions than SRS scores and therefore more likely to introduce noise in the association between underlying social skills measured at one point and response latencies measured over several sessions (Chatham et al., 2018).

To investigate how turn-by-turn information flow affects response timing (RQ4), we measured the predictability of utterances using a computational approach. For each pair of successive utterances, we calculated how similar or different they were from one another in terms of their semantic content (i.e., cosine similarity). This measure, based on language models (specifically, the all-mpnet-base-v2 model from the Sentence Transformers library (Reimers & Gurevych, 2020)), captures how predictable an utterance is given the previous one. Higher predictability scores indicate that the content of an utterance is more closely related to or follows more naturally from the previous turn. Lower predictability scores indicate novel information being introduced, unexpected responses, or even topic shifts and new conversational directions. By examining how these predictability scores relate to response latencies, we can assess whether children respond differently to more versus less predictable conversational turns, and whether they take longer to produce utterances that represent greater shifts in the conversation.

To investigate how interpersonal adjustment affects turn-taking (RQ5), we used a multivariate outcome vector autoregression, conditioning the response latencies of both the child and adult on each other's concurrent as well as lagged latencies, including conversational context as a moderator. Varying effects in the two equations were modelled as correlated between children and their interlocutors. This model is thus similar to structural equation modeling approaches such as the actor-partner interdependence (Cook & Kenny, 2005) and the dyadic coupling (Helm et al., 2012) models.

While our analyses focus primarily on children's turn-taking abilities across contexts, we acknowledge that these behaviors emerge within dyadic interactions where adult interlocutors significantly shape the conversation. In addition to research question 5, which focuses specifically on interpersonal adjustment between children and caregivers, complete parallel analyses of adult behaviors are provided in the **Supplementary Materials (Tables S1, S2, S3, S4 and Figures S1, S2, S3, S4)**, reflecting our recognition of turn-taking as fundamentally interpersonal and as an emergent property of the dyadic system.

Estimates from the models were reported as mean and 95% Credible Intervals (CI) of the posterior estimates. We calculated evidence ratios (ER) for our hypotheses in the form of the posterior probability of the directed hypothesis against the posterior probability of

all the alternatives; that is, if we expected higher response latencies in autistic children, we would count the posterior samples compatible with this hypothesis and divide them by the number of posterior samples compatible with a null or negative effect. An evidence ratio of 5 thus implies that the evidence in favour of the hypothesis is 5 times greater than the evidence in favour of the alternative model or hypothesis. An evidence ratio of 'Inf' (short for infinite) implies that all posterior samples are in the direction of the hypothesis, such that the ratio in favour becomes infinite. Throughout the results section, we compare posterior estimates between those in the experimenter and parent conversations, on the one hand, and between those in parent conversations and the matching game (conducted only with parents), on the other hand, in order to ascertain the effect of the familiarity of the interlocutor and the effect of experimental task on response latencies, respectively. Full details on model specification, including weakly informative priors, and model quality checks are available in the **Supplementary Materials**.

2.4.1 Control Analyses

To evaluate the impact of overlaps on response latency patterns for RQ1–RQ3, we conducted two control analyses. First, we re-ran the models excluding overlaps (i.e., negative latencies) to determine how the inclusion of overlaps in the main analyses potentially influenced our results. Except for specifying a truncated prior indicating no value could be below 0 to facilitate the sampling process, these control models and their specifications were identical to the primary models. Second, to assess the extent of overlap in each group for each task, we ran a multi-level logistic regression model, with the rate of overlap conditioned on diagnostic group, conversational context, and familiarity of the interlocutor, using the same varying effects structure as in the primary model.

To examine the impact of long pauses on group differences, we conducted a second control analysis. We fitted a simpler version of the primary model where the exponential rate parameter remained constant across contexts, while retaining all other model specifications. This allowed us to evaluate whether failing to account for varying rates of long pauses across contexts affected the estimation of group differences in response timing. Without modelling context-specific variation in long pauses, the model may overestimate response latencies for children who produced fewer long pauses than expected or underestimate latencies for children who produced more long pauses than expected.

These control analyses provide critical context for interpreting group differences, helping to distinguish the relative roles of overlaps and long pauses in the results and enable better interpretation of the findings in relation to the existing literature. Complete results of these analyses are available in the **Supplementary Materials**.

In addition to these analyses of overlaps and long pauses, we conducted four further control analyses. Firstly, for our model of information flow, we conducted a control analysis to check whether short, low-content utterances like backchannels unduly affected the analysis (Corps et al., 2022) and to make researcher degrees of freedom more explicit in our analyses. We included backchannels in our main analyses as they provide meaningful responsive contributions to conversation that affect the flow and timing of interactions, but acknowledge that different theoretical perspectives exist on their status. While one could argue backchannels provide cheap signals of common ground and do not require as much cognitive processing to respond

to (Dideriksen, Christiansen, Tylén, et al., 2023), we consider them an integral part of the informational landscape, requiring speakers to process contributions and identify appropriate turn transition points. We ran control models with utterances with fewer than three words excluded (see **Table S12**). Secondly, due to prior research showing gender differences in turn-taking dynamics (Cho et al., 2023; Pauline et al., 2024), we conducted an analysis of how response times changed across genders (see **Table S7**). Thirdly, because the familiarity of the experimenter increased with visits, we checked whether there were any changes in child response latencies according to this increase in familiarity (see **Figure S10**). Fourthly, as a final overall control analysis, we created surrogate dyads to emulate the baseline turn-taking that would happen by chance if there were no contingency between adult and child responses (see **Table S13**). The surrogate pairs were created by pairing the time codes of the caregiver of a different child within the same diagnostic group and visit with the time codes of the child. The aim of this control analysis was to show how our patterns of results could not be reduced to simple differences in vocalization rates and duration but were due to the specific contingency in the interpersonal temporal sequence of vocalizations. We also provide an additional Cross-Recurrence Quantification Analysis to quantify non-linear dynamical coupling patterns in dyadic interactions (e.g., Dale & Spivey (2006)), which offers a complementary perspective on how behavioral patterns between interaction partners synchronize, diverge, and recur across time (see Supplementary Materials).

2.5. Code availability

All code and data required to reproduce this research paper are publicly available and documented on OSF and GitHub.

3. RESULTS

3.1. Social Context and Diagnostic Group (RQ1 & RQ2)

3.1.1 Average Response Latencies (μ)

The results pertaining to RQ1 and RQ2 are shown in **Table 3** and **Figure 2**. There were strong differences across social contexts. Children exhibited slower response latencies when conversing with the unfamiliar experimenter than with their parent in both autism (186ms [2.33, 367.05], ER = 19.3) and typical development groups (353ms [113, 591], ER = 93.3). The children also produced slower latencies when solving the matching game than when more freely conversing with their parents in both the autism (279ms [124, 430], ER = 93.3) and typical development groups (297ms [131, 466], ER = 160.3). Diagnostic group also mattered. Response latencies for autistic children were faster than for those in the typical development group when aggregating across conditions (-136ms [-267, 0], ER = 18.7). This effect was the strongest in the conversations with the unfamiliar experimenter (-242ms [-476, -5], ER = 20.1), followed by the matching game (-93ms [-249, 67], ER = 5, difference in strength from the conversation with the experimenter: -167ms [-466, 133], ER = 4.6) and the conversations with the parent (-75ms [-272, 139], ER = 2.8, difference in strength from the matching game: -18ms [-254, 212], ER = 0.8). Test-retest reliability of response latency estimates across

the sessions was high within each social context, indicating stable individual differences in conversational timing over repeated sessions (see **Table 3**). Additionally, response latencies exhibited high cross-context consistency within each child, with those showing faster latencies in one social context tending to be faster in others (**Figure S5**). An analysis of how response times changed across genders showed a slight tendency for girls and non-binary children to be even faster than boys in the autism group: comparisons between the autism and typical development group, however, remained robust (see **Table S7**). There were also no clear developmental changes in child response latencies according to the increasing familiarity of the experimenter over the course of the multiple visits (see **Figure S10**).

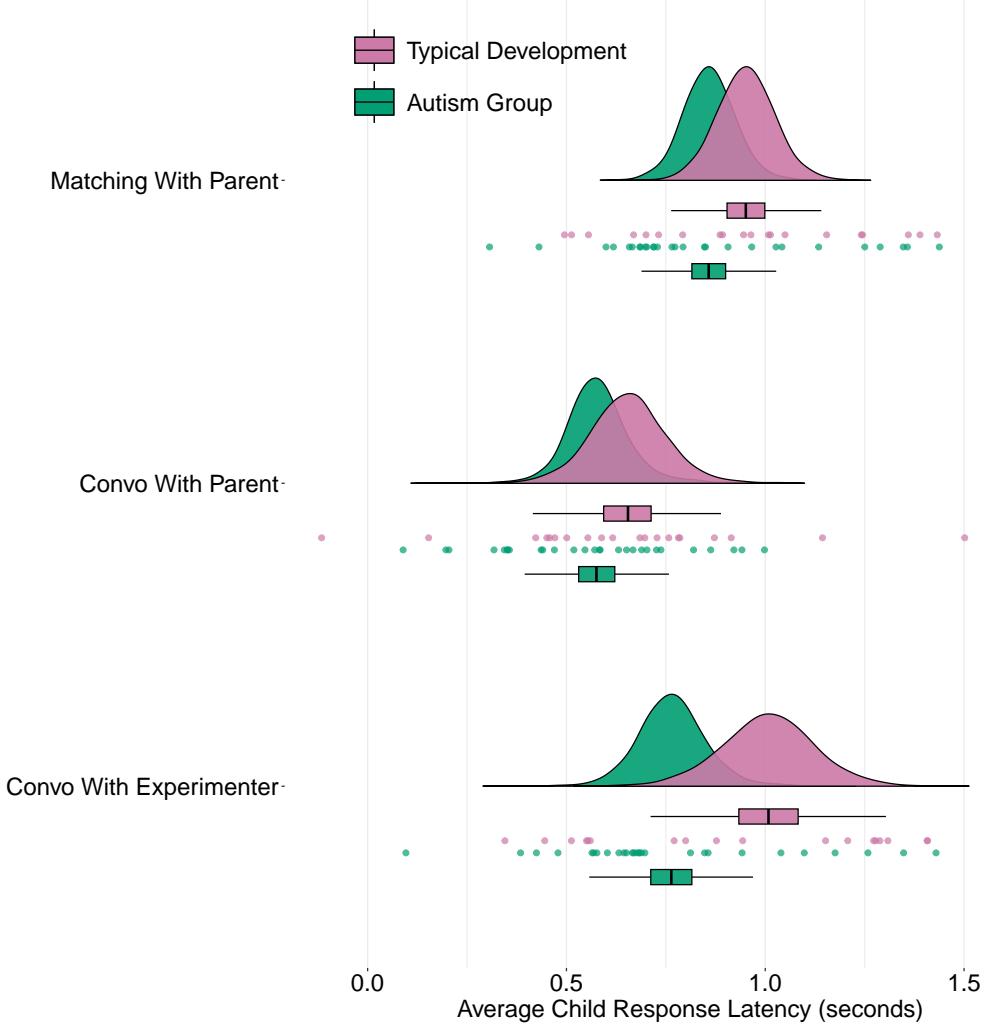


Figure 2: Model estimates for child response latencies across different conversational contexts for autistic children (green) and typically developing children (purple). The points denote participant-level posterior predictions from the model. The density plots and boxplots show aggregated posterior predictions for the three different social contexts: Matching Game, Convo With Parent (Conversations with Parents) and Convo With Experimenter (Conversations with Experimenter). The boxplots show the median (vertical line dividing the box), interquartile range (box edges) and most extreme data points (whiskers) for the aggregated posterior predictions. This figure shows the estimated μ parameter (central tendency) from our ex-Gaussian model and specifically captures the location of the main distribution after accounting for systematic differences in extended pauses (right tails), which are separately modelled by the beta parameter. Note that these values differ from raw data averages as they account for the hierarchical data structure and separate the distributional components.

Table 3: Posterior estimates for child response latencies across individual conditions (*Matching Game*, *Convo With Parent* (*Conversations with Parents*) and *Convo With Experimenter* (*Conversations with Experimenter*)) and aggregated across conditions (*Aggregate*) for different model parameters: the Gaussian component (*Mean Latency*) in milliseconds, proportion of latencies below zero (*Overlap Proportion*) in proportions, variance (*Sigma*) in milliseconds, exponential component (*Beta*, or long pauses) on a log scale, and test-retest reliability across sessions in correlations. All of the parameter estimates listed in this table have an evidence ratio above 40 for a test of difference to null.

	Task	Autism Group	Typical Development
Mean Latency	Matching With Parent	859ms [753, 967]	952ms [833, 1069]
	Convo With Parent	580ms [460, 716]	655ms [497, 814]
	Convo With Experimenter	766ms [630, 908]	1007ms [810, 1199]
	Aggregate Estimate	735ms [651, 822]	871ms [767, 974]
Overlap Proportion	Matching With Parent	0.22 [0.19, 0.26]	0.22 [0.19, 0.25]
	Convo With Parent	0.23 [0.19, 0.28]	0.17 [0.13, 0.22]
	Convo With Experimenter	0.24 [0.2, 0.29]	0.2 [0.15, 0.24]
	Aggregate Estimate	0.23 [0.21, 0.26]	0.19 [0.17, 0.22]
Sigma	Matching With Parent	618ms [572, 668]	529ms [482, 578]
	Convo With Parent	541ms [485, 604]	601ms [523, 691]
	Convo With Experimenter	954ms [882, 1030]	898ms [826, 975]
	Aggregate Estimate	683ms [649, 720]	658ms [619, 701]
Beta	Matching With Parent	0.19 [0.12, 0.27]	0.28 [0.18, 0.38]
	Convo With Parent	-0.35 [-0.48, -0.23]	-0.31 [-0.45, -0.17]
	Convo With Experimenter	-0.26 [-0.37, -0.14]	-0.18 [-0.33, -0.03]
	Aggregate Estimate	-0.14 [-0.2, -0.08]	-0.07 [-0.15, 0.01]
Test-Retest Reliability	Matching With Parent	0.42 [0.27, 0.53]	0.42 [0.24, 0.65]
	Convo With Parent	0.64 [0.44, 0.77]	0.62 [0.35, 0.8]
	Convo With Experimenter	0.36 [0.05, 0.59]	0.66 [0.52, 0.8]

3.1.2 Frequency and Impact of Overlaps

The proportion of overlaps changed across social contexts and diagnostic group (see **Table 3**). Autistic children were more likely to overlap with adults compared to typically developing children when aggregated across contexts (0.04 [0, 0.08], ER = 21.1). Conversations with parents displayed the biggest group difference (0.06 [0, 0.12], ER = 15.8) followed by conversations with the experimenter (0.05 [-0.01, 0.11], ER = 10.4, strength of difference to conversations with the parent: -0.01 [-0.09, 0.07], ER = 1.5) and the matching game (0.01 [-0.04, 0.05], ER = 1.6, only slightly smaller strength of difference compared to conversations with parents: -0.05 [-0.12, 0.01], ER = 9.2). Removing overlaps did not substantially affect the overall findings on response latencies; the autism group was still faster than the typical development group in the matching game and conversations with the experimenter, albeit not with parents (see **Table S5** and **Figure S6**).

3.1.3 Variance in Response Latencies (σ)

Different social contexts were related to a different degree of variance in response latencies (see **Table 3**). Conversations with an unfamiliar experimenter displayed the highest variance for both the autism group (on log scale: 0.57 [0.43, 0.7], ER = Inf) and the typical development group (0.4 [0.23, 0.57], ER = Inf), followed by the matching game for the autism group (comparison to conversation with parent: 0.13 [0, 0.27], ER = 18.6) and conversation with parent for the typical development group (comparison to matching game: -0.13 [-0.29, 0.03], ER = 9.2). Autistic children displayed a slightly higher degree of variance than typically developing children when aggregated across conditions (on log scale: 0.04 [-0.04, 0.12], ER = 3.5). The strongest between-group difference was in the matching game (on log scale: 0.16 [0.04, 0.28], ER = 72.5) followed by the conversations with parents (on log scale: -0.1 [-0.28, 0.07], ER = 5.1, difference to conversations with parents on log scale: 0.26 [0.05, 0.48], ER = 40.3) and conversations with the experimenter (on log scale: 0.06 [-0.06, 0.17], ER = 4.3, difference to parent conversations: on log scale: 0.16 [-0.05, 0.38], ER = 9).

3.1.4 Long Pauses (β)

The prevalence of long pauses (i.e., the exponential tail) was influenced by social context and diagnostic group (see **Table 3**). The matching game displayed a higher tendency for long pauses than the conversations with the unfamiliar experimenter in both the autism group (0.55 [0.42, 0.69], ER = Inf) and typical development group (0.59 [0.42, 0.76], ER = Inf). Children in both the autism group and typical development group had a slightly lower tendency to include long pauses when conversing with parents than when conversing with the experimenter (Autism Group: 0.1 [-0.07, 0.27], ER = 5.5; Typical Development Group: 0.13 [-0.06, 0.33], ER = 6.7). Autistic children produced fewer long pauses when aggregated across conditions (on a log scale: -0.07 [-0.17, 0.03], ER = 6.5) and during both the matching game (-0.08 [-0.21, 0.04], ER = 6.8) and experimenter conversations (-0.08 [-0.26, 0.11], ER = 3.2), but not during the conversations with parents (-0.04 [-0.23, 0.15], ER = 1.9). There were no reliable differences in the strength of group differences across social contexts.

3.2. Individual Differences (RQ3)

The results pertaining to RQ3 are shown in **Table 4** and **Figure 3**. Estimates denote the change in child response latencies as a function of one standard deviation increase in each measure. Our control analyses excluding overlaps displayed analogous findings across all measures of individual differences, as shown in **Table S10** and **Figure S8**.

3.2.1 Social Cognition

Children with higher social cognition skills displayed faster responses across groups and contexts, although the effect was much more evident in typically developing children (Autism Group: -65ms [-185, 57], ER = 4.3; Typical Development: -257ms [-592, 79], ER = 8.4; Difference: 192ms [-171, 551], ER = 4.3). The influence of social cognition changed according to social context. The response latencies of typically developing children - but not of autistic children - when compared to the conversations with parents showed a bigger relation to social

cognition in the conversation with the unfamiliar experimenter (difference: -486ms [-1253, 300], ER = 5.8) as well as in the matching game (-309ms [-902, 267], ER = 4.4).

Table 4: *Posterior estimates for how child response latencies change across individual conditions (Matching Game, Conversations With Parents and Conversations With Experimenter) and all conditions (Aggregate Estimate) as a function of each of the types of skills listed below (Motor, Cognitive Skills, Language Skills, Social Awareness and Social Motivation). The three top measures (i.e., Social Cognition, Social Awareness, Social Motivation) came from the SRS-2 scale, Language skills from CELF-5 scale, and Motor skills from VAB-3. The estimates denote the change in child response latency as a function of one standard deviation increase in the skills of the children.*

Skill	Task	Autism Group	Typical Development
Social Cognition	Matching With Parent	-89ms [-233, 52]	-301ms [-690, 114]
	Convo With Parent	-78ms [-238, 84]	8ms [-511, 554]
	Convo With Experimenter	-27ms [-234, 179]	-478ms [-1098, 169]
	Aggregate Estimate	-65ms [-185, 57]	-257ms [-592, 79]
Social Awareness	Matching With Parent	19ms [-127, 168]	228ms [-62, 517]
	Convo With Parent	52ms [-108, 216]	29ms [-354, 412]
	Convo With Experimenter	-19ms [-232, 198]	452ms [-10, 910]
	Aggregate Estimate	18ms [-107, 146]	236ms [-14, 486]
Social Motivation	Matching With Parent	65ms [-63, 193]	-105ms [-424, 218]
	Convo With Parent	17ms [-124, 157]	117ms [-320, 567]
	Convo With Experimenter	82ms [-102, 268]	248ms [-269, 773]
	Aggregate Estimate	55ms [-53, 164]	87ms [-187, 370]
Language	Matching With Parent	35ms [-42, 113]	-6ms [-113, 102]
	Convo With Parent	21ms [-68, 112]	10ms [-140, 163]
	Convo With Experimenter	72ms [-41, 185]	45ms [-140, 231]
	Aggregate Estimate	43ms [-23, 110]	17ms [-80, 116]
Motor	Matching With Parent	4ms [-79, 88]	-26ms [-180, 128]
	Convo With Parent	-54ms [-152, 45]	-40ms [-251, 175]
	Convo With Experimenter	-19ms [-141, 101]	-19ms [-278, 252]
	Aggregate Estimate	-23ms [-96, 51]	-29ms [-169, 113]

3.2.2 Social Awareness

Typically developing children - but not autistic children - displayed a strong association between social awareness and responses across contexts (Autism Group: 18ms [-107, 146], ER = 1.4; Typical Development: 236ms [-14, 486], ER = 16.1; difference: -219ms [-498, 59], ER = 9.4). The effect of social awareness for the typical development group was most evident in the conversation with the experimenter compared to in the matching game (423ms [-131, 976], ER = 8.6) as well as in the matching game compared to the conversation with parents (199ms [-224, 624], ER = 3.7).

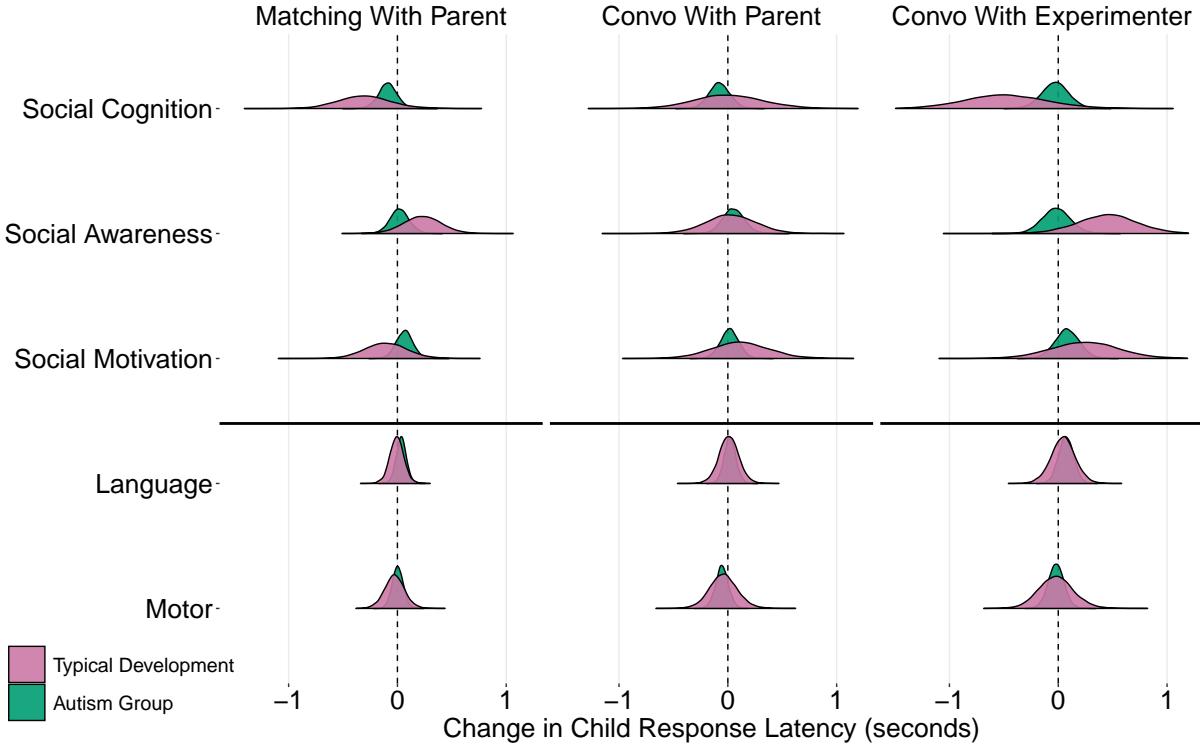


Figure 3: Model estimates for change in child response latencies as a function of one standard deviation increase in each measure across the three social contexts: Matching Game, Convo With Parent (Conversations with Parents) and Convo With Experimenter (Conversations with Experimenter). The boxplots show the median (vertical line deviding the box), interquartile range (box edges) and most extreme datapoints (whiskers) for the aggregated posterior predictions. The three top measures (i.e., Social Cognition, Social Awareness, Social Motivation) come from the SRS-2, whereas Language Skills refer to CELF-5 and Motor Skills refer to VAB-3.

3.2.3 Social Motivation

The relation between social motivation and response latencies was small and unsystematic across groups and contexts. Autistic children with higher social motivation showed a consistent but only very slight slowdown (55ms [-53, 164], ER = 4), especially in interactions with the experimenter (82ms [-102, 268], ER = 3.4). Typically developing children only displayed the same effect in the conversation with an unfamiliar experimenter (248ms [-269, 773], ER = 3.7) and the opposite effect in the matching game with parents (-105ms [-424, 218], ER = 2.4).

3.2.4 Language Skills

Similarly, there was only anecdotal evidence for a relation between language skills and response latencies across groups and contexts. We restrict the reporting here to the strongest absolute effect and direct the reader to **Table 4** to see the other effects. Autistic children with higher language skills showed an overall small tendency for slower response times compared to the typical development group, but there was no clear evidence for a difference between

groups (autism group: 43ms [-23, 110], ER = 5.8; typical development: 17ms [-80, 116], ER = 1.6, difference: 26ms [-92, 144], ER = 1.8).

3.2.5 Motor Skills

Lastly, there was likewise only anecdotal evidence for a relation between motor skills and response latencies across groups and contexts, and we direct the reader to **Table 4** to see other effects not reported here. Children with higher motor skills had slightly faster response latencies in the autism group (-23ms [-96, 51], ER = 2.4) as well as in the typical development group (-29ms [-169, 113], ER = 1.8), but there was no evidence for a difference between groups (5ms [-156, 165], ER = 1.1).

3.3. Predictability (RQ4)

The results pertaining to RQ4 are shown in **Table 5** and **Figure 4**. Our control analyses excluding overlaps and short utterances displayed equivalent findings, as shown in **Table S12**.

3.3.1 Predictability of Previous Adult Utterance

There was strong evidence that both autistic and typically developing children responded faster to more predictable utterances from their interlocutor (Autism Group: -28ms [-48, -9], ER = 101; Typical Development: -32ms [-61, -3], ER = 25.2; difference: 4ms [-32, 40], ER = 1.3). Conversations with parents generated bigger effects of predictability across groups compared to those with the experimenter (Autism Group: -27ms [-80, 25], ER = 4; Typical Development: -61ms [-138, 19], ER = 9.1). There were weaker effects in the matching game compared to in conversations with the parent, but only for the typically developing children (Autism Group: -2ms [-48, 42], ER = 0.9; Typical Development: 53ms [-14, 121], ER = 9.3).

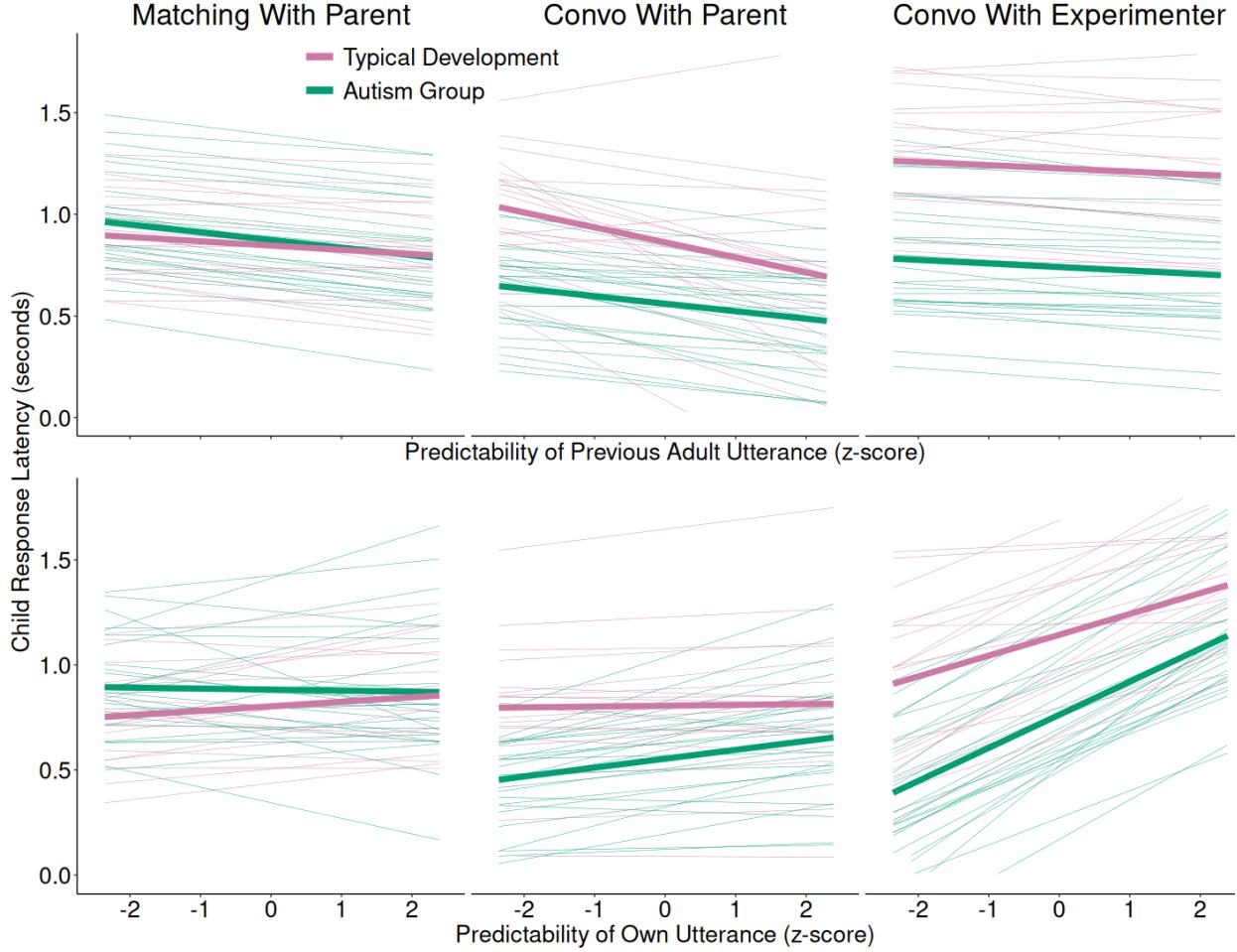


Figure 4: Model estimates for how child response latencies change as a function of the predictability of the previous adult response latency (top) and predictability of their own utterance (bottom) across the three social contexts: Matching Game, Convo With Parent (Conversations with Parents) and Convo With Experimenter (Conversations with Experimenter). The cosine similarity score here represents scaled values (i.e., one unit increase equals one standard deviation increase in cosine similarity). The faded lines are posterior predictions from the model for individual child participants in the study, whereas the thicker lines are average predictions.

3.3.2 Predictability of Own Utterance

There was strong evidence that both autistic and typically developing children responded faster if they were planning a low predictability utterance than a high predictability utterance (Autism Group: 69ms [47, 91], ER = Inf; Typical Development: 50ms [22, 79], ER = 383.6, difference: 18ms [-18, 54], ER = 4.1). The effects were more evident in conversations with the unfamiliar experimenter compared to with parents (Autism Group: -123ms [-181, -66], ER = 4999; Typical Development: -113ms [-186, -41], ER = 121), and lower in the matching game compared to conversations with parents for the autism group but not for the typical development group (Autism Group: -48ms [-99, 3], ER = 15.6; Typical Development: 19ms

[-35, 76], ER = 2.4).

Table 5: Posterior estimates for how child response latencies change across individual conditions (Matching Game, Parent Conversations and Experimenter Conversations) and all conditions (Aggregate Estimate) as a function of the predictability of the previous adult utterance and child utterance. The estimates denote the change in child response latency as a function of one standard deviation increase in the predictability of the utterance

	Tasks	Autism Group	Typical Development
Predictability of Previous Adult Utterance	Matching With Parent	-39ms [-65, -11]	-17ms [-53, 21]
	Convo With Parent	-37ms [-71, -1]	-70ms [-127, -11]
	Convo With Experimenter	-10ms [-48, 30]	-9ms [-63, 45]
	Aggregate Estimate	-28ms [-48, -9]	-32ms [-61, -3]
Predictability of Own Utterance	Matching With Parent	-4ms [-39, 32]	25ms [-14, 68]
	Convo With Parent	44ms [7, 80]	6ms [-30, 42]
	Convo With Experimenter	167ms [122, 211]	119ms [56, 182]
	Aggregate Estimate	69ms [47, 91]	50ms [22, 79]

3.4. Interpersonal Adjustment (RQ5)

The results pertaining to RQ5 are shown in **Table 6** and **Figure 5**. The estimates denote the change in child response latency as a function of a one-second increase in the response latency of the previous utterance. The analysis revealed a disparity between autistic and typically developing children in both interpersonal and self adjustment of response latencies across social contexts. When looking at the adult interlocutors, we see weak positive effects across different social contexts and no robust differences between the autism and typical development group (see **Figure S4** and **Table S4** in Supplementary Materials).

3.4.1 Interpersonal Adjustment

Autistic children demonstrated a small but reliable positive adjustment to their interlocutors' response latencies, whereas typically developing children did not (Autism Group: 38ms [16, 60], ER = 374; Typical Development: 10ms [-17, 35], ER = 2.8; difference: 29ms [-6, 63], ER = 11.1). Autistic children displayed a higher degree of adjustment in conversations with parents compared to in the matching game (difference: -40ms [-85, 1], ER = 17.4) as well as compared to in conversations with the experimenter (difference: -50ms [-118, 17], ER = 8.6). In comparison, typically developing children displayed no effect of the latency of the previous adult turn, and no clear differences across contexts.

Table 6: Posterior estimates for how child response latencies change across individual conditions (Matching Game, Parent Conversations and Experimenter Conversations) and all conditions (Aggregate Estimate) as a function of the latency of the previous adult utterance (Interpersonal Adjustment) and previous child response latency (Self Adjustment). The estimates denote the change in child response latency as a function of a one-second increase in the response latency of the previous utterance.

	Tasks	Autism Group	Typical Development
Interpersonal Adjustment	Matching With Parent	28ms [9, 47]	13ms [-14, 39]
	Convo With Parent	68ms [32, 107]	-6ms [-55, 40]
	Convo With Experimenter	18ms [-34, 71]	22ms [-31, 74]
	Aggregate Estimate	38ms [16, 60]	10ms [-17, 35]
Self Adjustment	Matching With Parent	24ms [7, 41]	31ms [7, 54]
	Convo With Parent	46ms [7, 89]	12ms [-36, 64]
	Convo With Experimenter	35ms [0.82, 69.39]	-5ms [-46, 38]
	Aggregate Estimate	35ms [17, 54]	13ms [-10, 36]

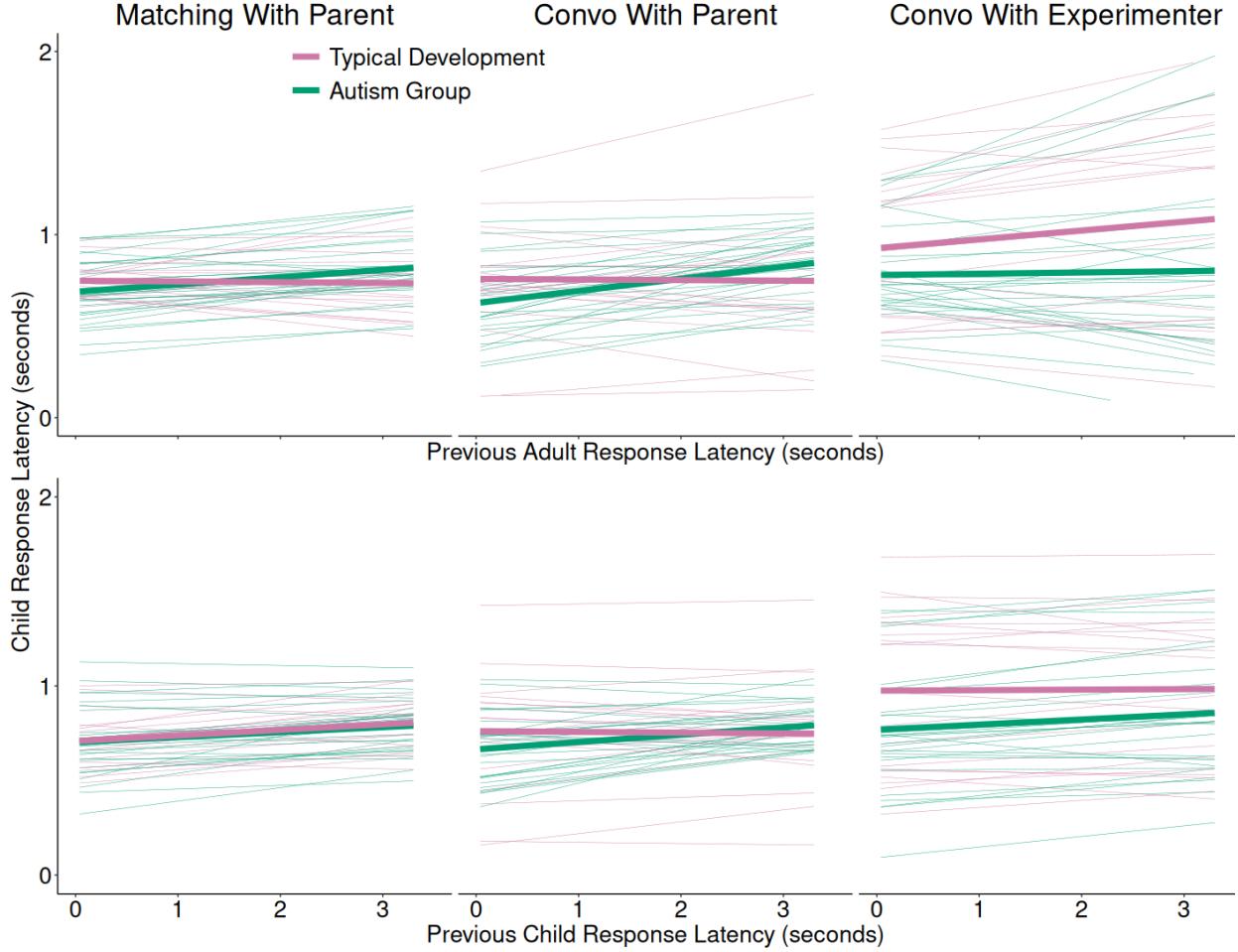


Figure 5: Model estimates for how child response latencies change with the latency of the previous adult response latency (top) and the previous child response latency (bottom) across the three social contexts: Matching Game, Convo With Parent (Conversations with Parents) and Convo With Experimenter (Conversations with Experimenter). The faded lines are posterior predictions from the model for individual child participants in the study, whereas the thicker lines are average predictions.

3.4.2 Self Adjustment

Children in both groups displayed a small positive adjustment to their own previous response latency, but the autistic children were more driven by their own previous response latency when aggregated across social contexts (Autism Group: 35ms [17, 54], ER = 749; Typical Development: 13ms [-10, 36], ER = 4.3; difference: 23ms [-8, 52], ER = 8.7). Children in the autism group adapted slightly less to their own response latencies in the matching game compared to when conversing with their parents (difference: -22ms [-66, 19], ER = 3.8), but there were no differences of this sort in the typical development group (difference: 19ms [-38, 72], ER = 2.7). The strongest difference between the groups was in conversations with the experimenter (difference: 40ms [-14, 94], ER = 7.7), followed by conversations with parents (difference: 34ms [-29, 99], ER = 4.4), and there were no differences in the matching game

(difference: -6ms [-35, 23], ER = 1.8).

4. DISCUSSION

In this study, we set out to challenge the idea of response latencies in turn-taking as a monolithic individual trait by investigating turn-taking as the interplay of individual differences, information flow and interpersonal dynamics – all of which are shaped by social context. We argue that a fuller understanding of child-adult turn-taking dynamics requires consideration of how results vary across multiple sessions and in different social contexts (RQ1), within atypical and typical development (RQ2) as well as how these dynamics interact with individual differences (RQ3) and turn-by-turn variations due to informational content (RQ4) and shared tempo (RQ5).

Our findings indicated that inter-turn response latencies were consistent across repeated measures *within social contexts*, but exhibited substantial differences across social contexts (RQ1). Conversations with a familiar interlocutor enabled faster turn-taking (i.e., faster child response latencies with a parent compared to an experimenter), and more structured tasks slowed down the conversation (i.e., slower responses in the matching game compared to conversations with their parent). We also found substantial differences between diagnostic groups (RQ2) modulated by social context; autistic children produced faster (more overlaps and shorter gaps) and more variable inter-turn response latencies compared to typically developing children – and these group differences were particularly salient in interactions with the unfamiliar interlocutor. Turn-taking dynamics were also modulated by individual differences among children (RQ3). Higher social cognition skills were related to faster response latencies in both groups, whereas children with higher social awareness and motivation exhibited slight slowdowns in responses in the typical development and autism group, respectively. Social context also shaped how turn-by-turn information flow related to response latencies (RQ4). The more predictable the interlocutor’s previous utterance and the less predictable the child’s own coming utterance, the faster the response. Yet, with unfamiliar experimenters, the predictability of the other’s utterance ceased to play a role, and the predictability of the child’s own utterance became a stronger predictor. Analogously, social context modulated the turn-taking dynamics of intra- and interpersonal adjustment (RQ5). Only the autism group adjusted their response times to those of the adult’s preceding turn – and only in interactions with their parents. Both groups displayed a weak but positive autocorrelation of their own successive response latencies, but less so when interacting with the experimenter.

Our findings demonstrate how deconstructing turn-taking into its constituent dimensions can reveal new insights on conversational dynamics. By explicitly including individual differences, moment-by-moment information flow, and interpersonal rhythms – and their relation to changes in social context – we identified how the interplay among these dimensions can shape inter-turn response latencies. This multidimensional approach proved particularly valuable in theorizing about two key phenomena: the faster response latencies of autistic children and the systematic effects of interlocutor familiarity. We view these timing patterns not as products of simple linear causation but as emergent properties that differentially scaffold interaction coherence across contexts. In the following sections, we consider individual and interpersonal mechanisms that may underlie the faster responses of autistic children

across social contexts. We then examine how interlocutor familiarity differentially shapes the interplay of group differences, individual characteristics, and turn-by-turn processes, providing a more nuanced understanding of its impact on turn-taking dynamics. Together, our approach reveals that turn-taking is a multicomponential construct that is shaped by individual differences, cognitive processing demands, and interpersonal adaptations – all of which are systematically modulated by the contextual demands of different conversational settings.

4.1. Why do we see Faster Responses in Autism?

Whereas earlier studies have found that autistic individuals exhibit slower response latencies (about 300ms) than neurotypicals (Ochi et al., 2019; Nguyen et al., 2022), we found that autistic children are faster by around 140ms (replicating R. Fusaroli et al. (2025), and to a certain extent Wehrle et al. (2023)). Understanding the source of this effect requires unpacking multiple factors. In conversations with parents, autistic children tended to produce more overlaps than neurotypical children, which in turn, may explain why they respond faster (see also R. Fusaroli et al. (2025)). However, even after excluding overlaps, autistic children still had shorter gaps in the matching game and in conversations with the experimenter (see **Table S5**). Autistic children also tended to produce fewer long pauses than typically developing children in the matching game; not modelling the long tail removes the group difference in the matching game, but not in the other conversations (see **Table S9**). Taken together, the findings observed in the conversation with parents and in the matching game could be discrepant from previous literature at least partially because of the differences in the statistical treatment of the data. Our findings for interactions with the unfamiliar interlocutor, on the other hand, continue to yield a clear group difference even when removing overlaps and long pauses. We argue that the explanation for this remaining difference is to be found in the underlying mechanisms of turn-taking. Crucially, conversations are both the expression of the individual interlocutors involved and of the interpersonal dynamics emerging between them. While autistic children might exhibit atypical processing and behaviours, adults might also bring different expectations to a conversation with autistic children and accordingly provide different scaffolding and reactions to them than to typically developing children (Ferjan Ramírez et al., 2020). Consequently, we discuss possible complementary explanations of group differences in terms of both child and adult behaviours in the interaction (R. Fusaroli et al., 2023a,b).

4.1.1 Individual Mechanisms: Child Executive Function & Impulsivity

Autistic children are often attributed a large suite of differences in socio-cognitive and other abilities (Frith & Happé, 1994; Fletcher-Watson & Happé, 2019). In our analyses, we included individual differences in social cognition, motivation and awareness, language and motor skills, which we had hypothesized to be directly involved in the development of turn-taking skills (RQ2). Nonetheless, group differences remained, even when stratifying by those individual differences. A complementary mechanism generating the prevalence of overlaps, short gaps and scarcity of long pauses could be located in the atypical executive function sometimes observed in autism (Russell, 1997; Hlavatá et al., 2018), and the perhaps related high co-

morbidity with ADHD and its impulsivity (Carrascosa-Romero & De Cabo-De La Vega, 2015; Bougeard et al., 2021; M. Fusaroli et al., 2023). Atypical executive function is often expressed as impulsivity and lack of inhibition (Hughes et al., 1994; Hill, 2004; Hlavatá et al., 2018; Geurts et al., 2014). For instance, as an utterance unfolds, each new word becomes more and more predictable. A child with high impulsivity might have a lower threshold for predictability, be more likely to lose attention, or just be generally impatient and jump in earlier. This might involve increased backchannels, or interrupting and neglecting the social nicety of letting the other finish (Dobbinson et al., 1998; Cola et al., 2022). Such a tendency to overlap might be further emphasized in the relaxed atmosphere of a conversation with a caregiver, where indeed the propensity to overlap is particularly high for autistic children.

4.1.2 Interpersonal Mechanisms: Adult Expectations & Predictability

Group differences in children's response latencies may also be driven by the behaviours of the adult in the interaction, and the feedback cycles of adjusting to each other (R. Fusaroli, Rączaszek-Leonardi, & Tylén, 2014; R. Fusaroli et al., 2019; Dingemanse et al., 2023; Cox, Bergmann, et al., 2023). For example, while adult interlocutors did display comparable response latencies, overlap proportions, long pauses and degrees of variability across the two groups (cf., **Table S1 and S6**), they also produced speech with a higher average turn-level predictability when interacting with autistic children (see **Table S11**). This is in line with previous work showing that parents of autistic children produce shorter utterances and employ a smaller vocabulary (R. Fusaroli et al., 2019). Thus, autistic children might not be more impulsive and more likely to interrupt their interlocutor; they may simply be reacting to predictable information in a typical way – and if typically developing children were to engage with the same predictable language, they may behave in similar ways to the autistic children¹. Differences in adult behaviours might be due to different expectations about their children on the basis of diagnostic status and individual differences - but also due to other attested differences in children's behaviours - e.g. atypical prosodic, verbal or multimodal patterns (R. Fusaroli et al., 2019; R. Fusaroli, Lambrechts, et al., 2017; R. Fusaroli et al., 2022; Rybner et al., 2022). Further, adults might be sensitive to interventions emphasizing how they need to produce good examples of social interactions and well-composed language (Ferjan Ramírez et al., 2020; Huber et al., 2023). Thus, a typically developing child overlapping with the adult might be quickly given the floor; while an autistic child might not, and the adult might keep speaking until the utterance reaches a possible conclusion.

4.2. Turn-Taking as Shaped by Interlocutor Familiarity

The familiarity of the interlocutor (i.e., caregiver versus experimenter) played a crucial role in shaping not only the children's response latencies, but also how differences by diagnostic group, individual characteristics and turn-by-turn dynamics played out; more so than the conversational task did. Engaging in verbal exchanges with unfamiliar interlocutors fundamentally differs from doing so with familiar interlocutors. For example, college students

¹Note that there is no reliable difference in reactions to predictability between groups (cf. Section 3.3.1), but that analysis focuses on turn-level predictability, and here we are talking about within-turn, word-level predictability.

responded faster when interacting with unfamiliar interlocutors to communicate active listening and engagement (Templeton et al., 2022, 2023). The children in this study, on the contrary, slowed their response latencies (by around 200ms in autism and 350ms in typical development), varied them more, and produced long pauses more frequently when interacting with the unfamiliar experimenter than with the familiar parent.

One possible explanation is that interactions with unfamiliar interlocutors are much more challenging for children. While adults can be prompted by the increased social pressure to give at least the impression of a prompt and smooth turn-taking, this might be too tall an order for children. Unfamiliarity brings the challenge of additional unpredictability and uncertainty to the interaction. The interlocutors do not share a history of interactional routines, and that makes inference and anticipation more challenging (Kimhi et al., 2014; Ying Sng et al., 2018; Pickering & Garrod, 2021; Tylén et al., 2013; De Marchena & Eigsti, 2016). We observe, for example, that the average predictability of the speech content in adult utterances was lower with unfamiliar than in familiar interlocutors (see **Table S11**), and this measure of predictability does not even account for the accumulated common ground and the multi-modal and social cues that children might have learned to pick up on in familiar adults throughout years of interactions.

The challenges involved in increased uncertainty might also be reflected in the relative lack of responsiveness to the utterance content of the unfamiliar interlocutor. The predictability of the other's utterances and previous response latencies mattered less when interacting with unfamiliar interlocutors - instead, response latencies were relatively more affected by the predictability of their own production. With increased uncertainty, children seem less able to promptly engage and pick up on cues. The finding that children responded faster when planning less predictable utterances may similarly suggest an information-theoretic mechanism: when preparing to contribute novel or unexpected information (low predictability), there may be greater communicative urgency to quickly take the floor. This urgency appears heightened in unfamiliar contexts, where establishing informational common ground becomes particularly important. Conversely, when planning more predictable responses that merely continue or affirm the previous speaker's contribution, the reduced information value may decrease the perceived pressure to respond rapidly.

In conversations with unfamiliar experimenters, we see the largest difference between autistic and typically developing children across social contexts, with typically developing but not autistic children showing a big relative slowdown. This is in line with complementary evidence that autistic children do not display as large a reaction as typically developing children in the presence of unfamiliar individuals compared to familiar ones (Stefanatos & Baron, 2011). For example, in contrast to typically developing children, autistic children show no difference in gaze aversion when interacting with familiar and unfamiliar interlocutors (Doherty-Sneddon et al., 2013), no difference in brain activation in response to familiar versus unfamiliar face stimuli (Dawson et al., 2002), and less flexibility in their behavioural repertoire across different social tasks (Forgeot d'Arc et al., 2020; Nadig et al., 2015).

Further, socio-cognitive skills seem to have a stronger impact on response latencies exactly when facing unfamiliar interlocutors. In typical development, the higher the socio-cognitive skills and the lower the social awareness, the faster the response latencies - both due to increased overlaps (cf., **Figure S8**) and shorter gaps (cf., **Table S10**). Social awareness might make the children more cautious and careful in the presence of a stranger, and perhaps

make them worried of cutting them off or just slower, while socio-cognitive skills might make them better at anticipating when an utterance is tapering to a conclusion. However, we did not see the same patterns in autistic children. Higher social motivation did slow autistic children down through fewer overlaps and longer gaps (cf., **Figure S8**) – perhaps again in cautious consideration of the unfamiliar situation, and them trying to do their best at steering clear from the interlocutor’s utterances – but higher socio-cognitive skills did not speed them up.

The marked diagnostic difference in the impact of socio-cognitive skills could be caused by many diverse mechanisms. Autistic children tend to have lower socio-cognitive skills (see **Table 1**) and thus might not be able to fully take advantage of social cues; or the measure of socio-cognitive skills might be more sensitive to differences in the lower values (more represented in the neurotypical than in the autistic group). Alternatively, anxiety and/or impulse control issues - often associated with autism (Carrascosa-Romero & De Cabo-De La Vega, 2015; Bougeard et al., 2021) - might make them less sensitive to social cues they might otherwise catch to various degrees in more familiar settings. Additionally, different conversational contexts impose varying psycholinguistic processing demands.

In sum, familiarity seems to have a strong impact on turn-taking and interact with individual and group differences as well as turn-by-turn dynamics. As unfamiliarity emphasizes differences by diagnosis and socio-cognitive skills, it can be considered a highly informative context to assess children’s turn-taking skills. However, the findings also emphasize that atypical performance with unfamiliar conversation partners only tell part of the story about children’s ability to engage with their everyday social context.

4.3. Limitations and Outlook

Since child populations, and particularly autistic children, are highly heterogeneous across different ages, much work still needs to be done to assess the generalizability of our work. For instance, younger (2-5 y.o.) children show similar but not identical patterns (e.g., with linguistic skills being better predictors of response latencies than in the current study (R. Fusaroli et al., 2025)). Despite matching groups on age at the group level, developmental differences across the relatively wide age range (6-17 years for the autism group and 6-15 years for typical development) of our participants might influence turn-taking dynamics in ways not fully captured by our analyses. Future work with larger samples could investigate how developmental trajectories of turn-taking differ across diagnostic groups and how they interact with social context demands.

As we start measuring more fine-grained individual differences, we focused on skills traditionally related to language acquisition and turn-taking: socio-cognitive, linguistic and motor skills. Yet, the current findings exhibited substantial heterogeneity in response latencies beyond differences in such skills (e.g. the remaining population-level differences between the autism and typical development groups). Network approaches to neurodevelopmental conditions highlight that individuals sharing the same diagnosis might present non-overlapping clinical profiles, and a better approach would be to rather fully decompose the diagnostic category into a wide range of interacting cognitive and clinical characteristics (Borsboom & Cramer, 2013; Fried & Cramer, 2017; Rocca et al., 2024), including measures of impulsivity. Further, measures of anxiety level could also be valuable, given the findings that unfamiliar

interlocutors elicited markedly different response latency patterns (Vasa et al., 2020). More work is needed in articulating how a wider range of individual differences - from sensory overload to information integration processes and attention (Pellicano & Burr, 2012; Murray et al., 2005; Simonsen et al., 2021) - might be involved in the mechanisms of turn-taking.

We also need to move towards a more interpersonal perspective on social interactions: how are adults shaping their behaviours and reactions when interacting with autistic children – and how are autistic children in turn reacting to this? A crucial component to achieving these steps will also be to integrate a detail-oriented qualitative analysis of what goes on in the utterances and at turn transitions (Bottema-Beutel et al., 2022; Sacks et al., 1974), including the social context, contingency and the diverse functions of gaps and overlaps. Besides generating novel research questions, qualitative analysis can be used to inform measurement and statistical modelling. One example is the measurement of utterance predictability. Whereas our measure was based purely on a short context of verbal data, utterance predictability is likely to be driven by several multimodal cues (e.g., gaze, prosody) as well as longer-term construction of common ground (Bögels & Torreira, 2021; Local & Walker, 2012; Schaffer, 1983). Our current measure of response latency is likewise agnostic as to whether the previous utterance has concluded or been interrupted or as to whether the current utterance is an answer, a completion, or a non-sequitur. Careful qualitative work has much to contribute to the understanding of the complexities and mechanisms of turn-taking (Bottema-Beutel et al., 2022).

5. CONCLUSION

Behind turn-taking lies a complex mutual interaction, intertwining contextual, individual and turn-by-turn processes (Harrist & Waugh, 2002; Schertz et al., 2018). Here we outline the central components underlying turn-taking: individual differences, properties of the utterance, and interpersonal dynamics, whose impact is deeply shaped by social context: the purpose of the conversation and the relation to the interlocutor. This framework construes turn-taking as immersed in a social context with different affordances for interactions. By considering these components, we not only need to reevaluate the group differences in turn-taking between neurotypical children and children with autism, but also the construct of turn-taking latency as a monolithic individual trait, or even indicator of social competence altogether. The picture we sketch of turn-taking is complex, but likely incomplete and more work is needed to develop this framework, including multimodality and multiple timescales of coordination.

6. CODE AND DATA AVAILABILITY STATEMENT

All code, data and materials required to reproduce this research paper are publicly available and documented on OSF and GitHub.

ACKNOWLEDGMENTS

We would like to acknowledge and thank all participating families for being involved with this research project, as well as James Fiumara, Monica Madya, Hannah Adeniyi, Jennifer Uzokwe. We would also like to thank audiences at the Max Planck Institute in Nijmegen, University of Melbourne, the International Society for Autism Research, and the Experimental Pragmatics Society for insightful feedback on preliminary results from this manuscript, as well as Andreas Hjorth Jeppesen for their helpful suggestions for improvement.

REFERENCES

- Berument, S. K., Rutter, M., Lord, C., Pickles, A., & Bailey, A. (1999). Autism screening questionnaire: diagnostic validity. *The British Journal of Psychiatry*, 175(5), 444–451.
- Bigelow, A. E. (1999). Infants' sensitivity to imperfect contingency in social interaction. *Early social cognition: Understanding others in the first months of life*, 137–154.
- Bögels, S., Magyari, L., & Levinson, S. C. (2015). Neural signatures of response planning occur midway through an incoming question in conversation. *Scientific reports*, 5(1), 12881.
- Bögels, S., & Torreira, F. (2021). Turn-end estimation in conversational turn-taking: The roles of context and prosody. *Discourse Processes*, 58(10), 903–924.
- Borsboom, D., & Cramer, A. O. (2013). Network analysis: an integrative approach to the structure of psychopathology. *Annual review of clinical psychology*, 9(1), 91–121.
- Bottema-Beutel, K., Crowley, S., & Kim, S. Y. (2022). Sequence organization of autistic children's play with caregivers: Rethinking follow-in directives. *Autism*, 26(5), 1267–1281.
- Bottema-Beutel, K., Lloyd, B., Watson, L., & Yoder, P. (2018). Bidirectional influences of caregiver utterances and supported joint engagement in children with and without autism spectrum disorder. *Autism Research*, 11(5), 755–765.
- Bougeard, C., Picarel-Blanchot, F., Schmid, R., Campbell, R., & Buitelaar, J. (2021). Prevalence of autism spectrum disorder and co-morbidities in children and adolescents: a systematic literature review. *Frontiers in psychiatry*, 12, 744709.
- Branigan, H. P., Bell, J., & McLean, J. F. (2016). Do you know what i know? the impact of participant role in children's referential communication. *Frontiers in psychology*, 7, 213.
- Bruni, T. P. (2014). Test review: Social responsiveness scale—second edition (srs-2). *Journal of Psychoeducational Assessment*, 32(4), 365–369. doi: 10.1177/0734282913517525
- Bryant, G. A. (2020). Evolution, structure, and functions of human laughter. In K. Floyd & R. Weber (Eds.), *The handbook of communication science and biology* (pp. 63–77). Routledge.
- Bryant, G. A., Fessler, D. M., Fusaroli, R., Clint, E., Aarøe, L., Apicella, C. L., ... Zhou, Y. (2016). Detecting affiliation in colaugher across 24 societies. *Proceedings of the National Academy of Sciences*, 113(17), 4682–4687.
- Bürkner, P.-C. (2017). brms: An r package for bayesian multilevel models using stan. *Journal of statistical software*, 80, 1–28.
- Carrascosa-Romero, M. C., & De Cabo-De La Vega, C. (2015). The comorbidity of adhd and autism spectrum disorders (asds) in community preschoolers. *ADHD-New Directions in Diagnosis and Treatment*, 109–162.

- Casillas, M., & Frank, M. C. (2017). The development of children's ability to track and predict turn structure in conversation. *Journal of Memory and Language*, 92, 234–253.
- Chatham, C., Taylor, K., Charman, T., Liogier D'Ardhuy, X., Eule, E., Fedele, A., ... del Valle Rubido, M. (2018). Adaptive behavior in autism: minimal clinically important differences on the vineland-II. *Autism Research*, 11(2), 270–283.
- Chesnut, S. R., Wei, T., Barnard-Brak, L., & Richman, D. M. (2017). A meta-analysis of the social communication questionnaire: Screening for autism spectrum disorder. *Autism*, 21(8), 920–928.
- Cho, S., Cola, M., Knox, A., Pelella, M. R., Russell, A., Hauptmann, A., ... Schultz, R. T. (2023). Sex differences in the temporal dynamics of autistic children's natural conversations. *Molecular Autism*, 14(1), 13. doi: <https://doi.org/10.1186/s13229-023-00545-6>
- Clark, E. V., & Lindsey, K. L. (2015). Turn-taking: A case study of early gesture and word use in answering where and which questions. *Frontiers in psychology*, 6, 890.
- Clark, H. H. (1996). *Using language*. Cambridge university press.
- Cola, M., Zampella, C. J., Yankowitz, L. D., Plate, S., Petrulla, V., Tena, K., ... Parish-Morris, J. (2022). Conversational adaptation in children and teens with autism: Differences in talkativeness across contexts. *Autism Research*, 15(6), 1090–1108.
- Constantino, J. N. (2012). *Social responsiveness scale*. Western Psychological Services.
- Cook, W. L., & Kenny, D. A. (2005). The actor–partner interdependence model: A model of bidirectional effects in developmental studies. *International Journal of Behavioral Development*, 29(2), 101–109.
- Corps, R. E., Knudsen, B., & Meyer, A. S. (2022). Overrated gaps: Inter-speaker gaps provide limited information about the timing of turns in conversation. *Cognition*, 223, 105037.
- Cox, C., Bergmann, C., Fowler, E., Keren-Portnoy, T., Roepstorff, A., Bryant, G., & Fusaroli, R. (2023). A systematic review and bayesian meta-analysis of the acoustic features of infant-directed speech. *Nature Human Behaviour*, 7(1), 114–133.
- Cox, C., Templeton, E., & Fusaroli, R. (2023). Fine-tuning social timing: From non-human to human animals and back. a commentary on the evolution of social timing by verga, kotz and ravignani (2023). *Physics of Life Reviews*, 47, 79–81.
- Cummins, F. (2019). *The ground from which we speak: Joint speech and the collective subject*. Cambridge Scholars Publishing.
- Dale, R. (2015). An integrative research strategy for exploring synergies in natural language performance. *Ecological Psychology*, 27(3), 190–201.
- Dale, R., Fusaroli, R., Duran, N. D., & Richardson, D. C. (2013). The self-organization of human interaction. In *Psychology of learning and motivation* (Vol. 59, pp. 43–95). Elsevier.

- Dale, R., & Spivey, M. J. (2006). Unraveling the dyad: Using recurrence analysis to explore patterns of syntactic coordination between children and caregivers in conversation. *Language Learning*, 56(3), 391–430.
- Dawson, G., Carver, L., Meltzoff, A. N., Panagiotides, H., McPartland, J., & Webb, S. J. (2002). Neural correlates of face and object recognition in young children with autism spectrum disorder, developmental delay, and typical development. *Child Development*, 73(3), 700–717.
- De Marchena, A., & Eigsti, I.-M. (2016). The art of common ground: Emergence of a complex pragmatic language skill in adolescents with autism spectrum disorders. *Journal of Child Language*, 43(1), 43–80.
- Denman, D., Speyer, R., Munro, N., Pearce, W. M., Chen, Y.-W., & Cordier, R. (2017). Psychometric properties of language assessments for children aged 4–12 years: A systematic review. *Frontiers in Psychology*, 8, 1515. doi: 10.3389/fpsyg.2017.01515
- Dideriksen, C. R., Christiansen, M. H., Dingemanse, M., Højmark-Bertelsen, M., Johansson, C., Tylén, K., & Fusaroli, R. (2023). Language-specific constraints on conversation: Evidence from danish and norwegian. *Cognitive Science*, 47(11), e13387.
- Dideriksen, C. R., Christiansen, M. H., Tylén, K., Dingemanse, M., & Fusaroli, R. (2023). Quantifying the interplay of conversational devices in building mutual understanding. *Journal of Experimental Psychology: General*, 152(3), 864.
- Dideriksen, C. R., Fusaroli, R., Tylén, K., Dingemanse, M., & Christiansen, M. H. (2019). Contextualizing conversational strategies: Backchannel, repair and linguistics alignment in spontaneous and task-oriented conversations. In *The 41st annual meeting of the cognitive science society* (pp. 261–267).
- Dilley, L., Lehet, M., Wieland, E. A., Arjmandi, M. K., Kondaurova, M., Wang, Y., ... Bergeson, T. (2020). Individual differences in mothers' spontaneous infant-directed speech predict language attainment in children with cochlear implants. *Journal of Speech, Language, and Hearing Research*, 63(7), 2453–2467.
- Dingemanse, M., & Liesenfeld, A. (2022). From text to talk: Harnessing conversational corpora for humane and diversity-aware language technology. In *Proceedings of the 60th annual meeting of the association for computational linguistics* (pp. 5614–5633).
- Dingemanse, M., Liesenfeld, A., Rasenberg, M., Albert, S., Ameka, F. K., Birhane, A., ... others (2023). Beyond single-mindedness: A figure-ground reversal for the cognitive sciences. *Cognitive Science*, 47(1), e13230.
- Dobbinson, S., Perkins, M. R., & Boucher, J. (1998). Structural patterns in conversations with a woman who has autism. *Journal of Communication Disorders*, 31(2), 113–134.
- Doherty-Sneddon, G., Whittle, L., & Riby, D. M. (2013). Gaze aversion during social style interactions in autism spectrum disorder and williams syndrome. *Research in developmental disabilities*, 34(1), 616–626.

- Dunn, J., & Shatz, M. (1989). Becoming a conversationalist despite (or because of) having an older sibling. *Child development*, 399–410.
- Ervin-Tripp, S. (1979). Children's verbal turn-taking. *Developmental pragmatics*, 391–414.
- Ervin-Tripp, S. (1982). Ask and it shall be given unto you: Children's requests. *Contemporary Perceptions of Language: Interdisciplinary Dimensions*, 235–245.
- Farmer, C., Adedipe, D., Bal, V., Chlebowski, C., & Thurm, A. (2020). Concordance of the vineland adaptive behavior scales, second and third editions. *Journal of intellectual disability research*, 64(1), 18–26.
- Fay, N., Walker, B., Swoboda, N., Umata, I., Fukaya, T., Katagiri, Y., & Garrod, S. (2018). Universal principles of human communication: Preliminary evidence from a cross-cultural communication game. *Cognitive Science*, 42(7), 2397–2413.
- Fay, W., & Schuler, A. L. (1980). *Emerging language in autistic children*. University Park Press.
- Ferjan Ramírez, N., Lytle, S. R., & Kuhl, P. K. (2020). Parent coaching increases conversational turns and advances infant language development. *Proceedings of the National Academy of Sciences*, 117(7), 3484–3491.
- Fletcher-Watson, S., & Happé, F. (2019). *Autism: A new introduction to psychological theory and current debate*. Routledge.
- Ford, C. E., Fox, B. A., & Thompson, S. A. (1996). Practices in the construction of turns: The “tcu” revisited. *Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA)*, 6(3), 427–454.
- Forgeot d'Arc, B., Devaine, M., & Daunizeau, J. (2020). Social behavioural adaptation in autism. *PLoS computational biology*, 16(3), e1007700.
- Fried, E. I., & Cramer, A. O. (2017). Moving forward: Challenges and directions for psychopathological network theory and methodology. *Perspectives on Psychological Science*, 12(6), 999–1020.
- Frith, U., & Happé, F. (1994). Autism: Beyond “theory of mind”. *Cognition*, 50(1-3), 115–132.
- Fusaroli, M., Pellegrini, L., Fusaroli, R., Raschi, E., Menchetti, M., & Poluzzi, E. (2023). Behavioral excess and disruptive conduct: A historical and taxonomic approach to the origin of the ‘impulse control disorders’ diagnostic construct. *Addiction*, 118(4), 763–770.
- Fusaroli, R., Cox, C. M. M., Weed, E., Szabó, B. I., Fein, D., & Naigles, L. (2025). The development of turn-taking skills in typical development and autism. *Cognitive Science*. doi: <https://doi.org/10.31234/osf.io/5ap6u>
- Fusaroli, R., Gangopadhyay, N., & Tylén, K. (2014). The dialogically extended mind: Language as skilful intersubjective engagement. *Cognitive Systems Research*, 29, 31–39.

- Fusaroli, R., Grossman, R., Bilenberg, N., Cantio, C., Jepsen, J. R. M., & Weed, E. (2022). Toward a cumulative science of vocal markers of autism: A cross-linguistic meta-analysis-based investigation of acoustic markers in american and danish autistic children. *Autism Research*, 15(4), 653–664.
- Fusaroli, R., Konvalinka, I., & Wallot, S. (2014). Analyzing social interactions: the promises and challenges of using cross recurrence quantification analysis. In *Translational recurrences: From mathematical theory to real-world applications* (pp. 137–155). Springer.
- Fusaroli, R., Lambrechts, A., Bang, D., Bowler, D. M., & Gaigg, S. B. (2017). Is voice a marker for autism spectrum disorder? a systematic review and meta-analysis. *Autism Research*, 10(3), 384–407.
- Fusaroli, R., Raczaszek-Leonardi, J., & Tylén, K. (2014). Dialog as interpersonal synergy. *New Ideas in Psychology*, 32, 147–157.
- Fusaroli, R., & Tylén, K. (2012). Carving language for social coordination: A dynamical approach. *Interaction studies*, 13(1), 103–124.
- Fusaroli, R., & Tylén, K. (2016). Investigating conversational dynamics: Interactive alignment, interpersonal synergy, and collective task performance. *Cognitive science*, 40(1), 145–171.
- Fusaroli, R., Tylén, K., Garly, K., Steensig, J., Christiansen, M., & Dingemanse, M. (2017). Measures and mechanisms of common ground: Backchannels, conversational repair, and interactive alignment in free and task-oriented social interactions. In *the 39th annual conference of the cognitive science society (cogsci 2017)* (pp. 2055–2060).
- Fusaroli, R., Weed, E., Fein, D., & Naigles, L. (2019). Hearing me hearing you: Reciprocal effects between child and parent language in autism and typical development. *Cognition*, 183, 1–18.
- Fusaroli, R., Weed, E., Rocca, R., Fein, D., & Naigles, L. (2023a). Caregiver linguistic alignment to autistic and typically developing children: A natural language processing approach illuminates the interactive components of language development. *Cognition*, 236, 105422.
- Fusaroli, R., Weed, E., Rocca, R., Fein, D., & Naigles, L. (2023b). Repeat after me? both children with and without autism commonly align their language with that of their caregivers. *Cognitive Science*, 47(11), e13369.
- Garvey, C., & Berninger, G. (1981). Timing and turn taking in children's conversations. *Discourse processes*, 4(1), 27–57.
- Gelman, A., Simpson, D., & Betancourt, M. (2017). The prior can often only be understood in the context of the likelihood. *Entropy*, 19(10), 555.
- Geurts, H. M., van den Bergh, S. F., & Ruzzano, L. (2014). Prepotent response inhibition and interference control in autism spectrum disorders: Two meta-analyses. *Autism Research*, 7(4), 407–420.

- Gratier, M., Devouche, E., Guellai, B., Infanti, R., Yilmaz, E., & Parlato-Oliveira, E. (2015). Early development of turn-taking in vocal interaction between mothers and infants. *Frontiers in psychology*, 6(1167), 236–245.
- Harrist, A. W., & Waugh, R. M. (2002). Dyadic synchrony: Its structure and function in children's development. *Developmental review*, 22(4), 555–592.
- Heeman, P. A., Lunsford, R., Selfridge, E., Black, L. M., & Van Santen, J. (2010). Autism and interactional aspects of dialogue. In *Proceedings of the SIGDIAL 2010 conference* (pp. 249–252).
- Heesen, R., & Fröhlich, M. (2022). *Revisiting the human ‘interaction engine’: comparative approaches to social action coordination* (Vol. 377) (No. 1859). The Royal Society.
- Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, 38(4), 555–568.
- Helm, J. L., Sbarra, D., & Ferrer, E. (2012). Assessing cross-partner associations in physiological responses via coupled oscillator models. *Emotion*, 12(4), 748.
- Hill, E. L. (2004). Executive dysfunction in autism. *Trends in cognitive sciences*, 8(1), 26–32.
- Hilton, C. B., Moser, C. J., Bertolo, M., Lee-Rubin, H., Amir, D., Bainbridge, C. M., ... others (2022). Acoustic regularities in infant-directed speech and song across cultures. *Nature Human Behaviour*, 6(11), 1545–1556.
- Hlavatá, P., Kašpárek, T., Linhartová, P., Ošlejšková, H., & Bareš, M. (2018). Autism, impulsivity and inhibition a review of the literature. *Basal Ganglia*, 14, 44–53.
- Huber, E., Ferjan Ramírez, N., Corrigan, N. M., & Kuhl, P. K. (2023). Parent coaching from 6 to 18 months improves child language outcomes through 30 months of age. *Developmental Science*, 26(6), e13391.
- Hughes, C., Russell, J., & Robbins, T. W. (1994). Evidence for executive dysfunction in autism. *Neuropsychologia*, 32(4), 477–492.
- Iverson, J. M. (2010). Developing language in a developing body: The relationship between motor development and language development. *Journal of child language*, 37(2), 229–261.
- Jaffe, J., Beebe, B., Feldstein, S., Crown, C. L., Jasnow, M. D., Rochat, P., & Stern, D. N. (2001). Rhythms of dialogue in infancy: Coordinated timing in development. *Monographs of the society for research in child development*, 66(2), 1–149.
- Jardat, P., Calandreau, L., Ferreira, V., Gouyet, C., Parias, C., Reignier, F., & Lansade, L. (2022). Pet-directed speech improves horses' attention toward humans. *Scientific Reports*, 12(1), 4297.
- Karasik, L. B., Tamis-LeMonda, C. S., & Adolph, K. E. (2014). Crawling and walking infants elicit different verbal responses from mothers. *Developmental science*, 17(3), 388–395.

- Kimhi, Y., Shoam-Kugelmas, D., Agam Ben-Artzi, G., Ben-Moshe, I., & Bauminger-Zviely, N. (2014). Theory of mind and executive function in preschoolers with typical development versus intellectually able preschoolers with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 44, 2341–2354.
- Kondaurova, M. V., Bergeson, T. R., & Xu, H. (2013). Age-related changes in prosodic features of maternal speech to prelingually deaf infants with cochlear implants. *Infancy*, 18(5), 825–848.
- Konvalinka, I., & Roepstorff, A. (2012). The two-brain approach: how can mutually interacting brains teach us something about social interaction? *Frontiers in Human Neuroscience*, 6, 215.
- Leezenbaum, N. B., Campbell, S. B., Butler, D., & Iverson, J. M. (2014). Maternal verbal responses to communication of infants at low and heightened risk of autism. *Autism*, 18(6), 694–703.
- Lemoine, N. P. (2019). Moving beyond noninformative priors: why and how to choose weakly informative priors in bayesian analyses. *Oikos*, 128(7), 912–928.
- Levinson, S. (2006). Cognition at the heart of human interaction. *Discourse studies*, 8(1), 85–93.
- Levinson, S. (2016). Turn-taking in human communication—origins and implications for language processing. *Trends in Cognitive Sciences*, 20(1), 6–14.
- Lindsay, L., Gambi, C., & Rabagliati, H. (2019). Preschoolers optimize the timing of their conversational turns through flexible coordination of language comprehension and production. *Psychological science*, 30(4), 504–515.
- Local, J., & Walker, G. (2012). How phonetic features project more talk. *Journal of the International Phonetic Association*, 42(3), 255–280.
- Magyari, L., Bastiaansen, M. C., De Ruiter, J. P., & Levinson, S. C. (2014). Early anticipation lies behind the speed of response in conversation. *Journal of Cognitive Neuroscience*, 26(11), 2530–2539.
- Magyari, L., & De Ruiter, J. P. (2012). Prediction of turn-ends based on anticipation of upcoming words. *Frontiers in Psychology*, 3, 376.
- McCrimmon, A. W., & Smith, A. D. (2013). Review of the wechsler abbreviated scale of intelligence, second edition (wasi-ii). *Autism*, 31(3), 337–341.
- Murray, D., Lesser, M., & Lawson, W. (2005). Attention, monotropism and the diagnostic criteria for autism. *Autism*, 9(2), 139–156.
- Nadig, A., Seth, S., & Sasson, M. (2015). Global similarities and multifaceted differences in the production of partner-specific referential pacts by adults with autism spectrum disorders. *Frontiers in Psychology*, 6, 1888.

- Neimy, H., Pelaez, M., Carrow, J., Monlux, K., & Tarbox, J. (2017). Infants at risk of autism and developmental disorders: Establishing early social skills. *Behavioral Development Bulletin*, 22(1), 6.
- Nguyen, V., Versyp, O., Cox, C., & Fusaroli, R. (2022). A systematic review and bayesian meta-analysis of the development of turn taking in adult–child vocal interactions. *Child Development*, 93(4), 1181–1200.
- O'Brien, A. M., Perrachione, T. K., Weil, L. W., Araujo, Y. S., Halverson, K., Harris, A., ... others (2023). Altered engagement of the speech motor network is associated with reduced phonological working memory in autism. *NeuroImage: Clinical*, 37, 103299.
- Ochi, K., Ono, N., Owada, K., Kojima, M., Kuroda, M., Sagayama, S., & Yamasue, H. (2019). Quantification of speech and synchrony in the conversation of adults with autism spectrum disorder. *PloS one*, 14(12), e0225377.
- Ochs, E., Kremer-Sadlik, T., Sirota, K. G., & Solomon, O. (2004). Autism and the social world: An anthropological perspective. *Discourse studies*, 6(2), 147–183.
- Okobi Jr, D. E., Banerjee, A., Matheson, A. M., Phelps, S. M., & Long, M. A. (2019). Motor cortical control of vocal interaction in neotropical singing mice. *Science*, 363(6430), 983–988.
- Olsen, K., & Tylén, K. (2023). On the social nature of abstraction: cognitive implications of interaction and diversity. *Philosophical Transactions of the Royal Society B*, 378(1870), 20210361.
- Panneton, R., Cristia, A., Taylor, C., & Christine, M. (2023). Positive valence contributes to hyperarticulation in maternal speech to infants and puppies. *Journal of Child Language*, 1–11.
- Parish-Morris, J., Pallathra, A. A., Ferguson, E., Maddox, B. B., Pomykacz, A., Perez, L. S., ... Brodkin, E. S. (2019). Adaptation to different communicative contexts: an eye tracking study of autistic adults. *Journal of neurodevelopmental disorders*, 11, 1–10.
- Pauline, M., Weyland, M., & Kissine, M. (2024). Structure and acoustics of the speech of verbal autistic preschoolers. *Journal of child language*, 51(3), 509–525.
- Pellicano, E., & Burr, D. (2012). When the world becomes ‘too real’: a bayesian explanation of autistic perception. *Trends in cognitive sciences*, 16(10), 504–510.
- Piazza, G., Martin, C. D., & Kalashnikova, M. (2022). The acoustic features and didactic function of foreigner-directed speech: A scoping review. *Journal of Speech, Language, and Hearing Research*, 65(8), 2896–2918.
- Pickering, M. J., & Garrod, S. (2021). *Understanding dialogue: Language use and social interaction*. Cambridge University Press.

- Pouw, W., & Holler, J. (2022). Timing in conversation is dynamically adjusted turn by turn in dyadic telephone conversations. *Cognition*, 222, 105015.
- Ravignani, A., & de Reus, K. (2019). Modelling animal interactive rhythms in communication. *Evolutionary Bioinformatics*, 15, 1176934318823558.
- Reimers, N., & Gurevych, I. (2020, 11). Making monolingual sentence embeddings multilingual using knowledge distillation. In *Proceedings of the 2020 conference on empirical methods in natural language processing*. Association for Computational Linguistics. Retrieved from <https://arxiv.org/abs/2004.09813>
- Ritwika, V., Pretzer, G. M., Mendoza, S., Shedd, C., Kello, C. T., Gopinathan, A., & Warlaumont, A. S. (2020). Exploratory dynamics of vocal foraging during infant-caregiver communication. *Scientific reports*, 10(1), 10469.
- Rocca, R., Pistilli, G., Maheshwari, K., & Fusaroli, R. (2024). Introducing ellips: An ethics-centered approach to research on llm-based inference of psychiatric conditions. In *Proceedings of the aaai/acm conference on ai, ethics, and society* (Vol. 7, pp. 1243–1254).
- Russell, J. E. (1997). *Autism as an executive disorder*. Oxford University Press.
- Rybner, A., Jessen, E. T., Mortensen, M. D., Larsen, S. N., Grossman, R., Bilenberg, N., ... others (2022). Vocal markers of autism: Assessing the generalizability of machine learning models. *Autism Research*, 15(6), 1018–1030.
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *language*, 50(4), 696–735.
- Schaffer, D. (1983). The role of intonation as a cue to turn taking in conversation. *Journal of phonetics*, 11(3), 243–257.
- Schertz, H. H., Odom, S. L., Baggett, K. M., & Sideris, J. H. (2018). Mediating parent learning to promote social communication for toddlers with autism: Effects from a randomized controlled trial. *Journal of autism and developmental disorders*, 48, 853–867.
- Simonsen, A., Fusaroli, R., Petersen, M. L., Vermillet, A.-Q., Bliksted, V., Mors, O., ... Campbell-Meiklejohn, D. (2021). Taking others into account: combining directly experienced and indirect information in schizophrenia. *Brain*, 144(5), 1603–1614.
- Sparrow, S. S., Cicchetti, D. V., & Saulnier, C. A. (2016). *Vineland adaptive behavior scales, third edition*. Pearson.
- Stefanatos, G. A., & Baron, I. S. (2011). The ontogenesis of language impairment in autism: a neuropsychological perspective. *Neuropsychology Review*, 21, 252–270.
- Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., ... Yoon, K.-E. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, 106(26), 10587–10592.

- Tamis-LeMonda, C. S., & Adolph, K. E. (2013). Social referencing in infant motor action. In B. D. Homer & C. S. Tamis-LeMonda (Eds.), *The development of social cognition and communication* (pp. 145–164). Psychology Press.
- Templeton, E. M., Chang, L. J., Reynolds, E. A., Cone LeBeaumont, M. D., & Wheatley, T. (2022). Fast response times signal social connection in conversation. *Proceedings of the National Academy of Sciences*, 119(4), e2116915119.
- Templeton, E. M., Chang, L. J., Reynolds, E. A., Cone LeBeaumont, M. D., & Wheatley, T. (2023). Long gaps between turns are awkward for strangers but not for friends. *Philosophical Transactions of the Royal Society B*, 378(1875), 20210471.
- Tylén, K., Fusaroli, R., Bundgaard, P. F., & Østergaard, S. (2013). Making sense together: A dynamical account of linguistic meaning-making. *Semiotica*, 2013(194), 39–62.
- Uther, M., Knoll, M. A., & Burnham, D. (2007). Do you speak e-ng-li-sh? a comparison of foreigner-and infant-directed speech. *Speech communication*, 49(1), 2–7.
- Vasa, R. A., Keefer, A., McDonald, R. G., Hunsche, M. C., & Kerns, C. M. (2020). A scoping review of anxiety in young children with autism spectrum disorder. *Autism Research*, 13(12), 2038–2057.
- Verga, L., Kotz, S. A., & Ravignani, A. (2023). The evolution of social timing. *Physics of Life Reviews*, 46, 131–151.
- Waade, P. T., Enevoldsen, K. C., Vermillet, A.-Q., Simonsen, A., & Fusaroli, R. (2023). Introducing tomsup: Theory of mind simulations using python. *Behavior Research Methods*, 55(5), 2197–2231.
- Warlaumont, A. S., Richards, J. A., Gilkerson, J., & Oller, D. K. (2014). A social feedback loop for speech development and its reduction in autism. *Psychological Science*, 25(7), 1314–1324.
- Wechsler, D. (2008). Wechsler adult intelligence scale–fourth edition (waits-iv). *APA PsycTests*. doi: <https://doi.org/10.1037/t15169-000>
- Wehrle, S., Cangemi, F., Janz, A., Vogeley, K., & Grice, M. (2023). Turn-timing in conversations between autistic adults: Typical short-gap transitions are preferred, but not achieved instantly. *Plos one*, 18(4), e0284029.
- West, K. L., & Iverson, J. M. (2021). Communication changes when infants begin to walk. *Developmental science*, 24(5), e13102.
- Wiig, E. H., Secord, W. A., & Semel, E. (2013). *Clinical evaluation of language fundamentals: CELF-5*. Pearson.
- Wilson, M., & Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychonomic bulletin & review*, 12, 957–968.

Ying Sng, C., Carter, M., & Stephenson, J. (2018). A systematic review of the comparative pragmatic differences in conversational skills of individuals with autism. *Autism & Developmental Language Impairments*, 3, 2396941518803806.

Zamm, A., Pfördresher, P. Q., & Palmer, C. (2015). Temporal coordination in joint music performance: effects of endogenous rhythms and auditory feedback. *Experimental Brain Research*, 233, 607–615.

S.1. SUPPLEMENTARY MATERIALS

Our Supplementary Materials include additional results on adult response latencies, test-retest reliability, control analyses as well as model information and quality checks:

S.1.1 Adult Results	2
S.1.1.1 Average Response Latencies	2
S.1.1.2 Individual Socio-Cognitive Differences, Adults	4
S.1.1.3 Predictability, Adults	6
S.1.1.4 Shared Tempo, Adults	8
S.1.2 Test-Retest Reliability	10
S.1.3 Control Analyses	12
S.1.3.1 Child Latency With No Overlaps	12
S.1.3.2 Adult Latency With No Overlaps	14
S.1.3.3 Child Latency Across Gender	16
S.1.3.4 Child Latency With No Sigma and No Beta	17
S.1.3.5 Child Latency With Sigma But No Beta	18
S.1.3.6 Child Overlaps according to Individual Differences	19
S.1.3.7 Child Individual Differences With No Overlaps	21
S.1.3.8 Predictability	22
S.1.3.9 Child Latency With Predictability, No Short Utterances	23
S.1.3.10 Child Latency With Increasing Familiarity	25
S.1.3.11 Child Latency For Surrogate Pairs	26
S.1.4 Model Information and Quality Checks	28
S.1.4.1 Choice of Priors	28
S.1.4.2 Prior and Posterior Predictive Checks	29
S.1.4.3 Prior-Posterior Update Plots	32
S.1.5 Time Code Reliability	36
S.1.6 Recurrence Analysis	38
S.1.7 Research Assistants	42
S.1.7.1 Assignment of Research Assistants	42
S.1.7.2 Effects of Research Assistant	43
S.1.8 Recording Quality Assessment	45

S.1.1. Adult Results

S.1.1.1 Average Response Latencies

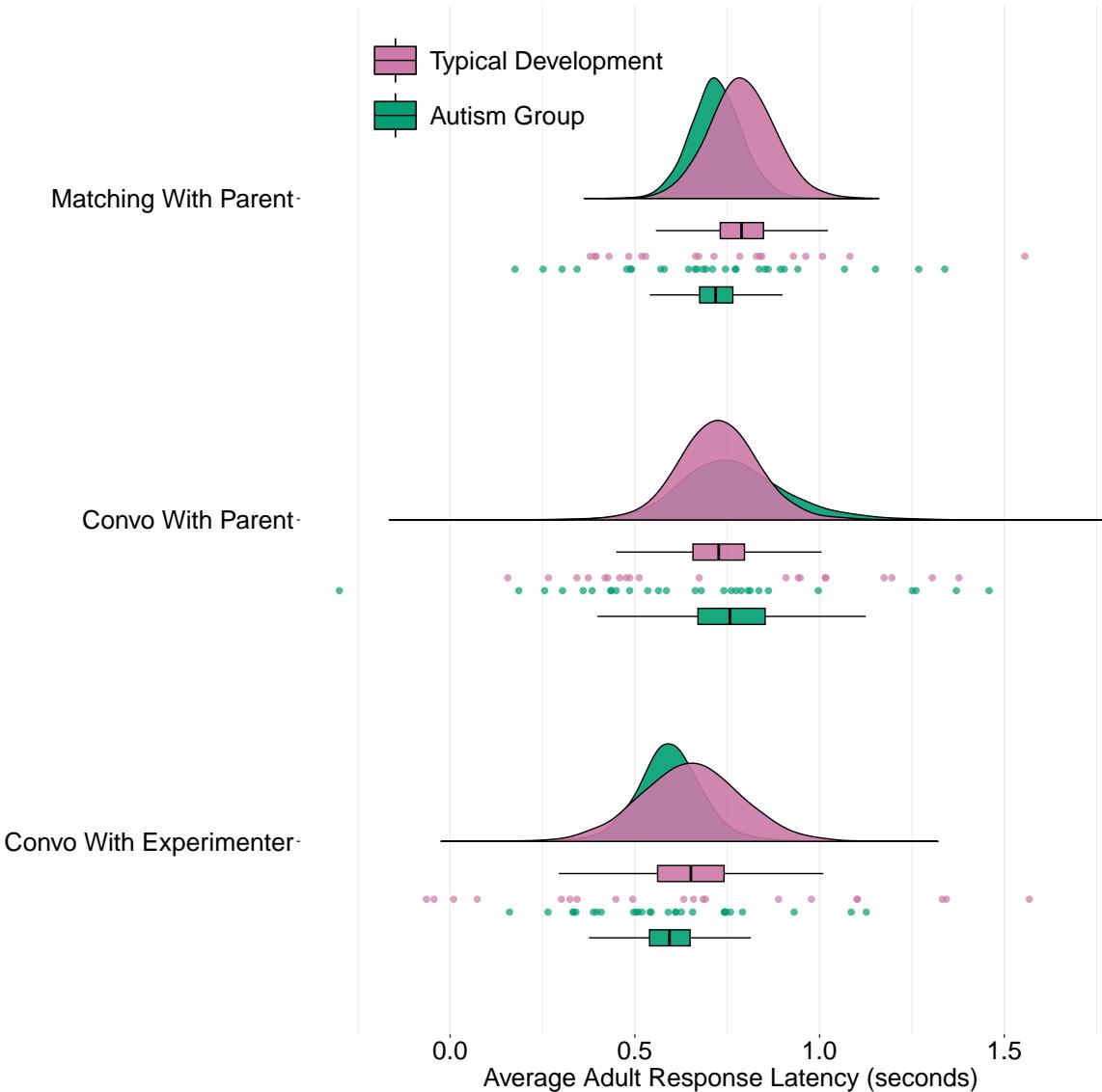


Figure S1: Model estimates for adult response latencies across different conversational contexts for autistic children (green) and typically developing children (purple). The points represent participant-level posterior predictions from the model. The density plots and boxplots show aggregated posterior predictions for the three different social contexts: Matching Game, Convo With Parent (Conversations with Parents) and Convo With Experimenter (Conversations with Experimenter). The boxplots show the median (vertical line deviding the box), interquartile range (box edges) and most extreme datapoints (whiskers) for the aggregated posterior predictions.

Table S1: Posterior estimates for adult response latencies across individual conditions (*Matching Game*, *Convo With Parent* (*Conversations with Parents*) and *Convo With Experimenter* (*Conversations with Experimenter*) and aggregated across conditions (*Aggregate Estimate*) for different model parameters: the Gaussian component (*Latency*) in milliseconds, proportion of latencies below zero (*Overlap Proportion*) in proportions, Standard Deviation (*Sigma*) in milliseconds, exponential component (*Beta*, or long pauses) on log scale, and *Test-Retest Reliability* across sessions in correlations.

	Tasks	Autism Group	Typical Development
Latency	Matching With Parent	720ms [606, 837]	790ms [645, 937]
	Convo With Parent	768ms [550, 1031]	728ms [552, 913]
	Convo With Experimenter	595ms [449, 743]	651ms [424, 874]
	Aggregate Estimate	694ms [590, 806]	723ms [612, 834]
Overlap Proportion	Matching With Parent	0.3 [0.26, 0.34]	0.26 [0.22, 0.3]
	Convo With Parent	0.22 [0.18, 0.27]	0.22 [0.17, 0.28]
	Convo With Experimenter	0.29 [0.25, 0.33]	0.26 [0.2, 0.33]
	Aggregate Estimate	0.27 [0.24, 0.3]	0.25 [0.22, 0.28]
Sigma	Matching With Parent	525ms [487, 566]	469ms [432, 510]
	Convo With Parent	499ms [439, 567]	568ms [492, 658]
	Convo With Experimenter	870ms [812, 933]	785ms [721, 855]
	Aggregate Estimate	611ms [579, 645]	594ms [558, 633]
Beta	Matching With Parent	0.19 [0.1, 0.27]	0.15 [0.05, 0.26]
	Convo With Parent	-0.14 [-0.28, 0]	-0.17 [-0.36, 0.01]
	Convo With Experimenter	-0.42 [-0.56, -0.28]	-0.33 [-0.49, -0.18]
	Aggregate Estimate	-0.12 [-0.2, -0.05]	-0.12 [-0.21, -0.02]
Test-Retest Reliability	Matching With Parent	0.47 [0.38, 0.56]	0.35 [0.08, 0.53]
	Parent Conversation	0.94 [0.93, 0.95]	0.82 [0.8, 0.86]
	Convo With Experimenter	0.78 [0.74, 0.8]	0.75 [0.72, 0.77]

S.1.1.2 Individual Socio-Cognitive Differences, Adults

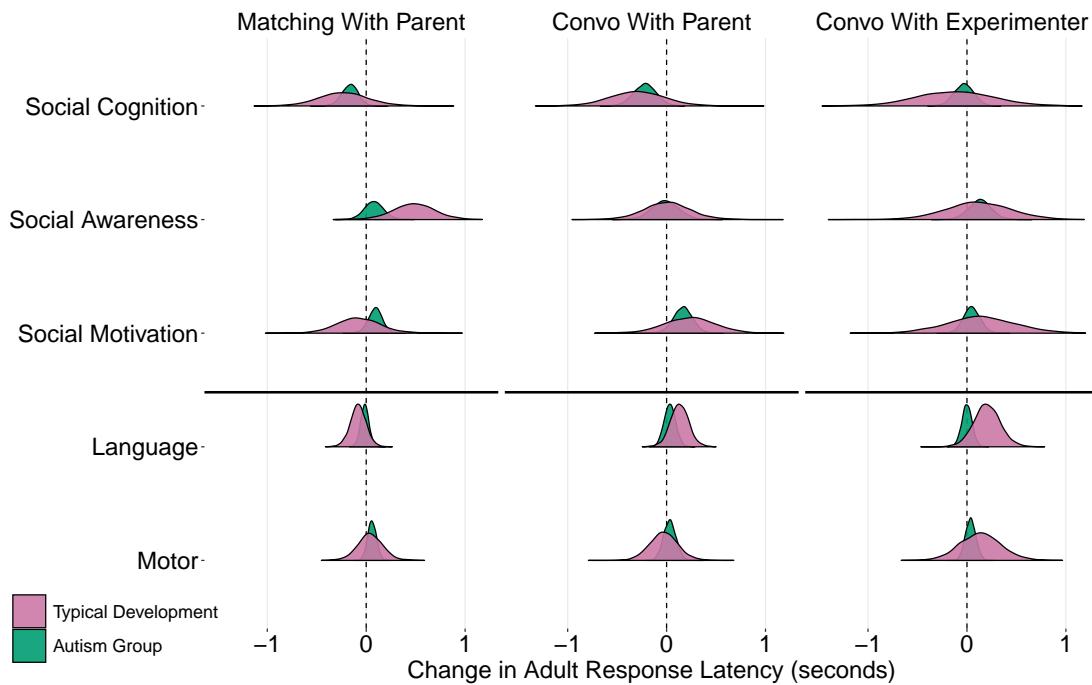


Figure S2: Model estimates for change in adult response latencies as a function of one standard deviation increase in each measure across the three social contexts: Matching Game, Convo With Parent (Conversations with Parents) and Convo With Experimenter (Conversations with Experimenter). The boxplots show the median (vertical line deviding the box), interquartile range (box edges) and most extreme datapoints (whiskers) for the aggregated posterior predictions. The three top measures (i.e., Social Cognition, Social Awareness, Social Motivation) come from the SRS-2, whereas Language Skills refer to CELF-5 and Motor Skills refer to VAB-3.

Table S2: Posterior estimates for how adult response latencies change across individual conditions (*Matching Game, Conversations With Parents and Conversations With Experimenter*) and all conditions (*Aggregate Estimate*) as a function of each of the types of skills listed below (*Motor, Cognitive Skills, Language Skills, Social Awareness and Social Motivation*). The three top measures (i.e., *Social Cognition, Social Awareness, Social Motivation*) came from the *Social Responsiveness Scale*, *Language skills* refer to the *Clinical Evaluation of Language Fundamentals*, and *Motor Skills* refer to *Vineland Adaptive Behaviour Scales*. The estimates denote the change in adult response latency as a function of one standard deviation increase in the skills of the children.

	Skills	Autism Group	Typical Development
Social Cognition	Matching With Parent	-165ms [-302, -33]	-228ms [-639, 192]
	Convo With Parent	-219ms [-407, -37]	-294ms [-751, 167]
	Convo With Experimenter	-31ms [-192, 132]	-118ms [-760, 538]
	Aggregate Estimate	-138ms [-232, -43]	-213ms [-518, 96]
Social Awareness	Matching With Parent	81ms [-77, 248]	483ms [127, 833]
	Convo With Parent	-12ms [-231, 213]	6ms [-385, 406]
	Convo With Experimenter	132ms [-55, 313]	112ms [-430, 662]
	Aggregate Estimate	67ms [-46, 182]	201ms [-51, 458]
Social Motivation	Matching With Parent	92ms [-29, 209]	-77ms [-437, 296]
	Convo With Parent	164ms [0.82, 323.53]	234ms [-194, 655]
	Convo With Experimenter	49ms [-88, 190]	145ms [-442, 737]
	Aggregate Estimate	102ms [20, 184]	101ms [-169, 378]
Language	Matching With Parent	-14ms [-84, 56]	-81ms [-217, 51]
	Convo With Parent	33ms [-63, 130]	131ms [-18, 285]
	Convo With Experimenter	-0.3ms [-86, 85]	204ms [-16, 426]
	Aggregate Estimate	6ms [-44, 57]	85ms [-16, 189]
Motor	Matching With Parent	60ms [-13, 133]	41ms [-171, 254]
	Parent Conversations	33ms [-71, 137]	-28ms [-262, 210]
	Convo With Experimenter	37ms [-48, 125]	137ms [-203, 485]
	Aggregate Estimate	43ms [-9, 96]	50ms [-104, 212]

S.1.1.3 Predictability, Adults

Table S3: *Posterior estimates for how adult response latencies change across individual conditions (Matching Game, Parent Conversations and Experimenter Conversations) and all conditions (Aggregate Estimate) as a function of the predictability of the previous child utterance and adult utterance. The estimates denote the change in adult response latency as a function of one standard deviation increase in the predictability of the utterance.*

	Tasks	Autism Group	Typical Development
Predictability of Previous Child Utterance	Matching With Parent	2.46ms [-26, 32]	37ms [6, 71]
	Convo With Parent	-10ms [-48, 28]	11ms [-29, 52]
	Convo With Experimenter	-36ms [-73, 1]	-34ms [-85, 16]
	Aggregate Estimate	-15ms [-35, 6]	4.66ms [-19, 29]
Predictability of Own Utterance	Matching With Parent	-8ms [-29, 14]	-4.16ms [-28, 19]
	Convo With Parent	-14ms [-38, 10]	4.98ms [-25, 35]
	Convo With Experimenter	47ms [14, 80]	55ms [16, 96]
	Aggregate Estimate	8ms [-8, 24]	19ms [-0.35, 37.7]

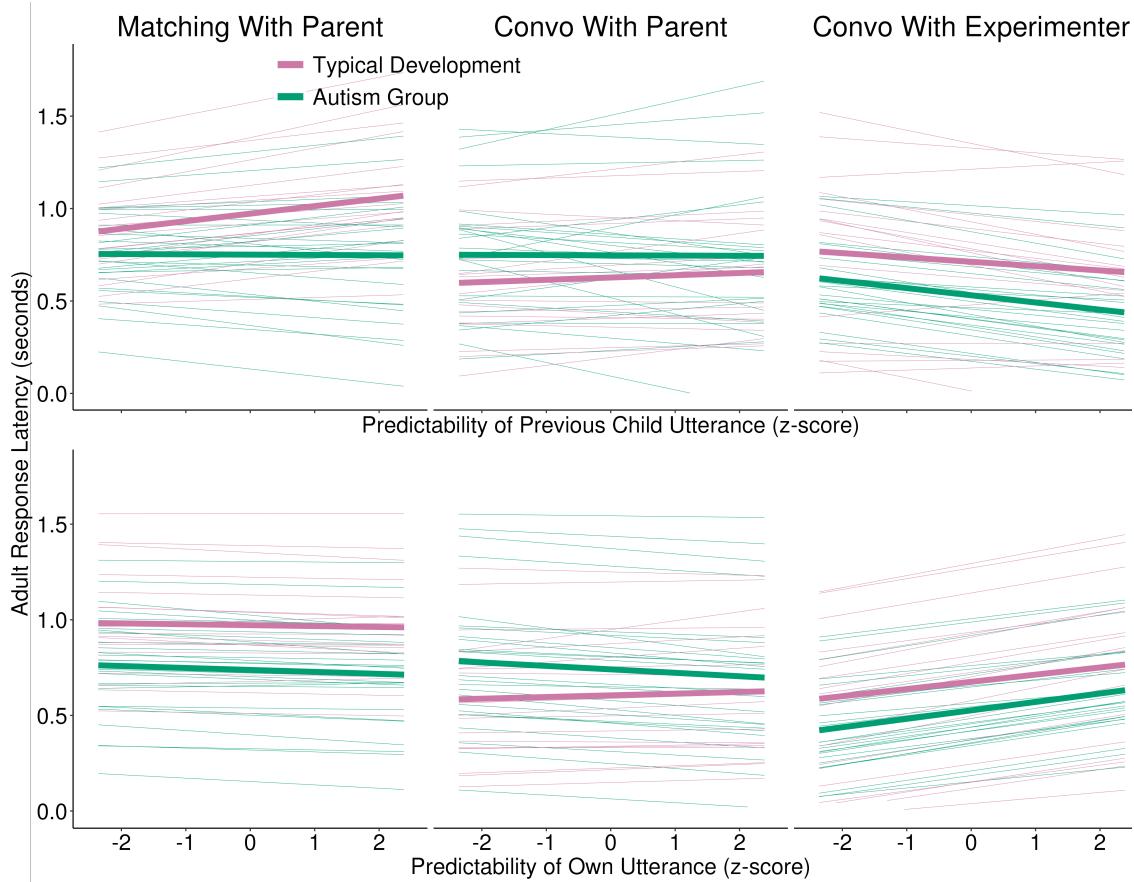


Figure S3: Model estimates for how adult response latencies change as a function of the predictability of the previous child response latency (top) and predictability of their own utterance (bottom) across the three social contexts: Matching Game, Convo With Parent (Conversations with Parents) and Convo With Experimenter (Conversations with Experimenter). The cosine similarity score here represents scaled values (i.e., one unit increase equals one standard deviation increase in cosine similarity). The faded lines are posterior predictions from the model for individual adult participants in the study, whereas the thicker lines are average predictions.

S.1.1.4 Shared Tempo, Adults

Table S4: Posterior estimates for how adult response latencies change across individual conditions (*Matching Game, Parent Conversations and Experimenter Conversations*) and all conditions (*Aggregate Estimate*) as a function of the latency of the previous child utterance (*Interpersonal Adjustment*) and previous adult response latency (*Self-Adjustment*). The estimates denote the change in adult response latency as a function of a one-second increase in the response latency of the previous utterance.

	Tasks	Autism Group	Typical Development
Interpersonal Adjustment	Matching With Parent	40ms [24, 58]	20ms [-8, 47]
	Convo With Parent	37ms [6, 70]	40ms [-12, 96]
	Convo With Experimenter	27ms [-10, 66]	45ms [0.85, 92.23]
	Aggregate Estimate	35ms [18, 53]	35ms [10, 61]
Self Adjustment	Matching With Parent	38ms [21, 55]	32ms [8, 58]
	Convo With Parent	67ms [38, 97]	46ms [10, 82]
	Convo With Experimenter	102ms [57, 148]	84ms [47, 121]
	Aggregate Estimate	69ms [50, 88]	54ms [35, 73]

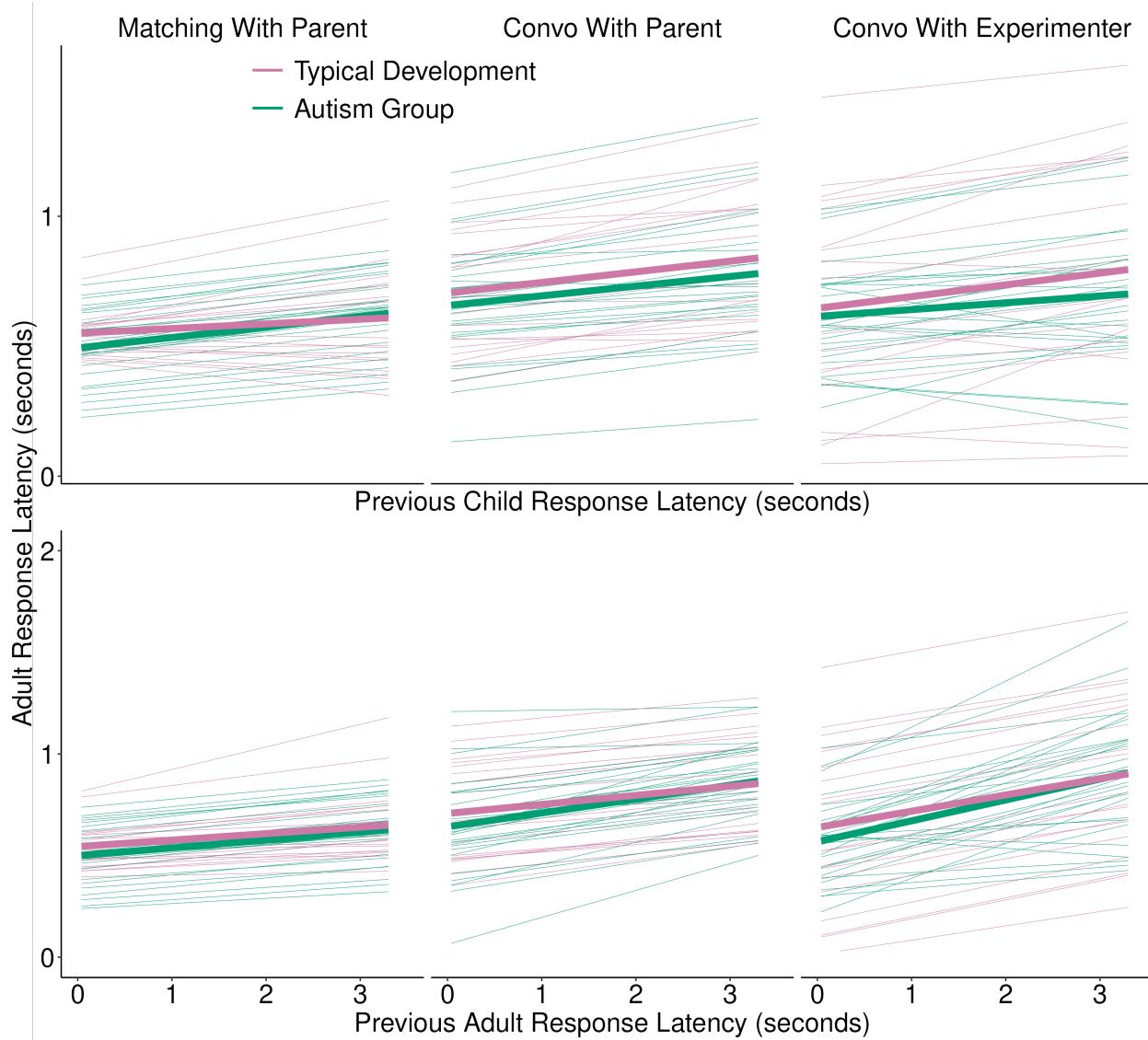


Figure S4: Model estimates for how adult response latencies change with the latency of the previous child response latency (top) and the previous adult response latency (bottom) across the three social contexts: Matching Game, Convo With Parent (Conversations with Parents) and Convo With Experimenter (Conversations with Experimenter). The faded lines are posterior predictions from the model for individual adult participants in the study, whereas the thicker lines are average predictions.

S.1.2. Test-Retest Reliability

To assess the consistency of individual children's response latencies across different social contexts, we calculated correlations between posterior estimates of each child's mean latency in each context. Cross-context consistency was quantified as the mean of all pairwise correlations between the three social contexts for each diagnostic group. Bootstrap confidence intervals ($n = 1000$) were computed to estimate uncertainty around these consistency measures. High cross-context consistency indicates that children who exhibit faster (or slower) response latencies in one social context tend to maintain this relative standing across other contexts, suggesting stable individual differences in conversational timing patterns. Cross-context consistency was moderate to good for both groups (Autism Group: $r = 0.65$, 95% CI [0.38, 0.82]; Typical Development: $r = 0.56$, 95% CI [0.21, 0.83]), indicating that individual differences in turn-taking latencies represent relatively stable characteristics of children's conversational behaviour across diverse social situations (see **Figure S5**).

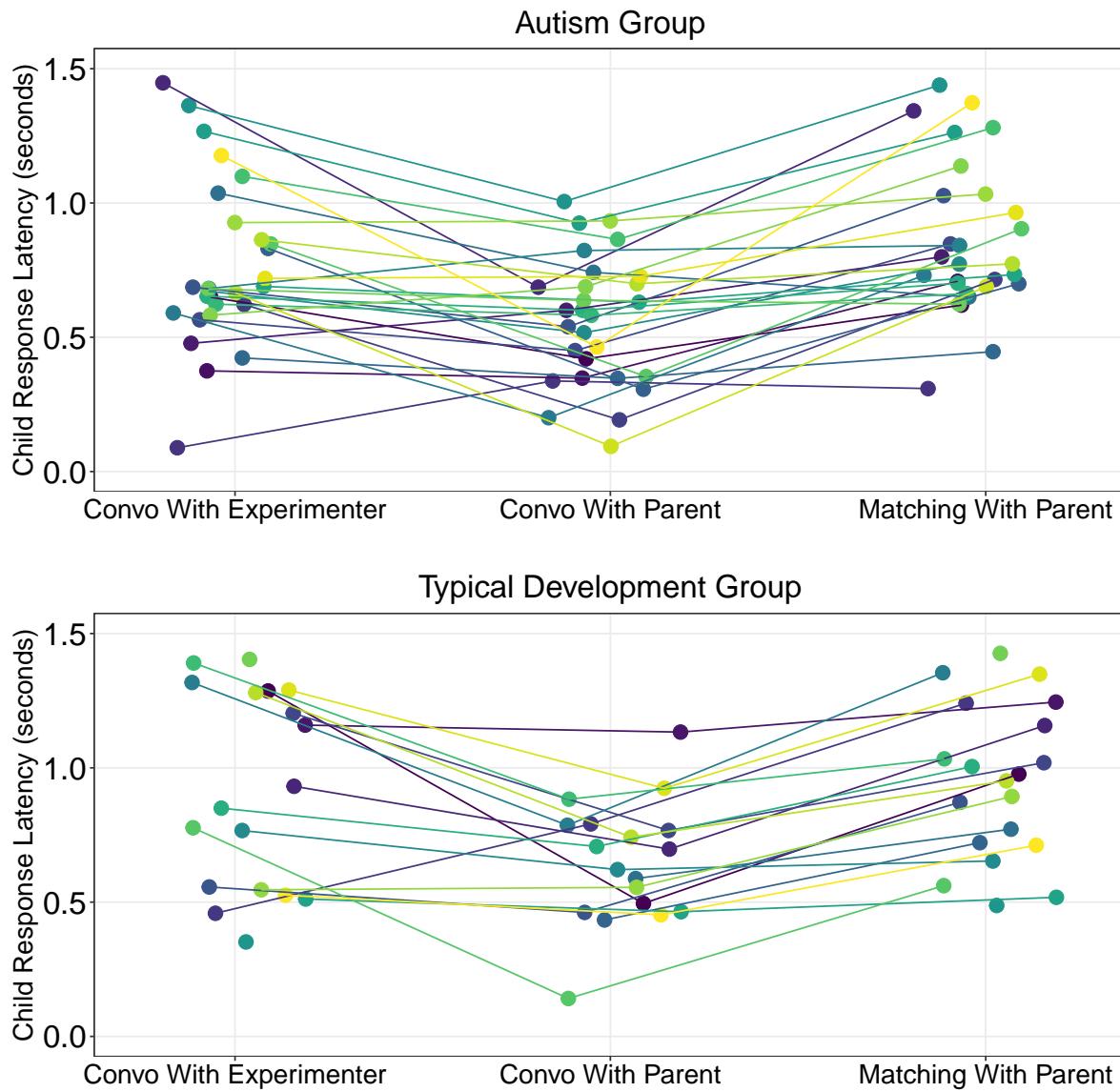


Figure S5: Panel of plots showing correlations among estimates of child-specific response latencies across different social contexts. Each point is the model estimate for a specific child and each child is connected across contexts with a line.

S.1.3. Control Analyses

S.1.3.1 Child Latency With No Overlaps

To check the extent to which the inclusion of overlaps influenced our estimates in the main manuscript, we ran control model without overlaps, the estimates for which are in **Table S5**.

Table S5: *Posterior estimates for child response latencies across individual conditions (Matching Game, Convo With Parent (Conversations with Parents) and Convo With Experimenter (Conversations with Experimenter) and aggregated across conditions (Aggregate Estimate) for different model parameters: the Gaussian component (Latency) in milliseconds, Standard Deviation (Sigma) in milliseconds, and exponential component (Beta, or long pauses) on log scale.*

	Tasks	Autism Group	Typical Development
Latency	Matching With Parent	1238ms [1181, 1298]	1328ms [1246, 1414]
	Convo With Parent	873ms [808, 938]	885ms [788, 986]
	Convo With Experimenter	1265ms [1165, 1360]	1454ms [1327, 1581]
	Aggregate Estimate	1125ms [1077, 1174]	1222ms [1152, 1291]
Beta	Matching With Parent	0.2 [0.15, 0.25]	0.27 [0.2, 0.33]
	Convo With Parent	-0.17 [-0.24, -0.09]	-0.17 [-0.27, -0.06]
	Convo With Experimenter	-0.09 [-0.17, -0.01]	-0.07 [-0.19, 0.05]
	Aggregate Estimate	-0.02 [-0.06, 0.03]	0.01 [-0.05, 0.07]
Sigma	Matching With Parent	7ms [3.76, 11.67]	6ms [3.33, 10.76]
	Convo With Parent	10ms [4.82, 20.03]	23ms [10, 45]
	Convo With Experimenter	201ms [163, 245]	326ms [286, 372]
	Aggregate Estimate	24ms [18, 33]	36ms [25, 48]

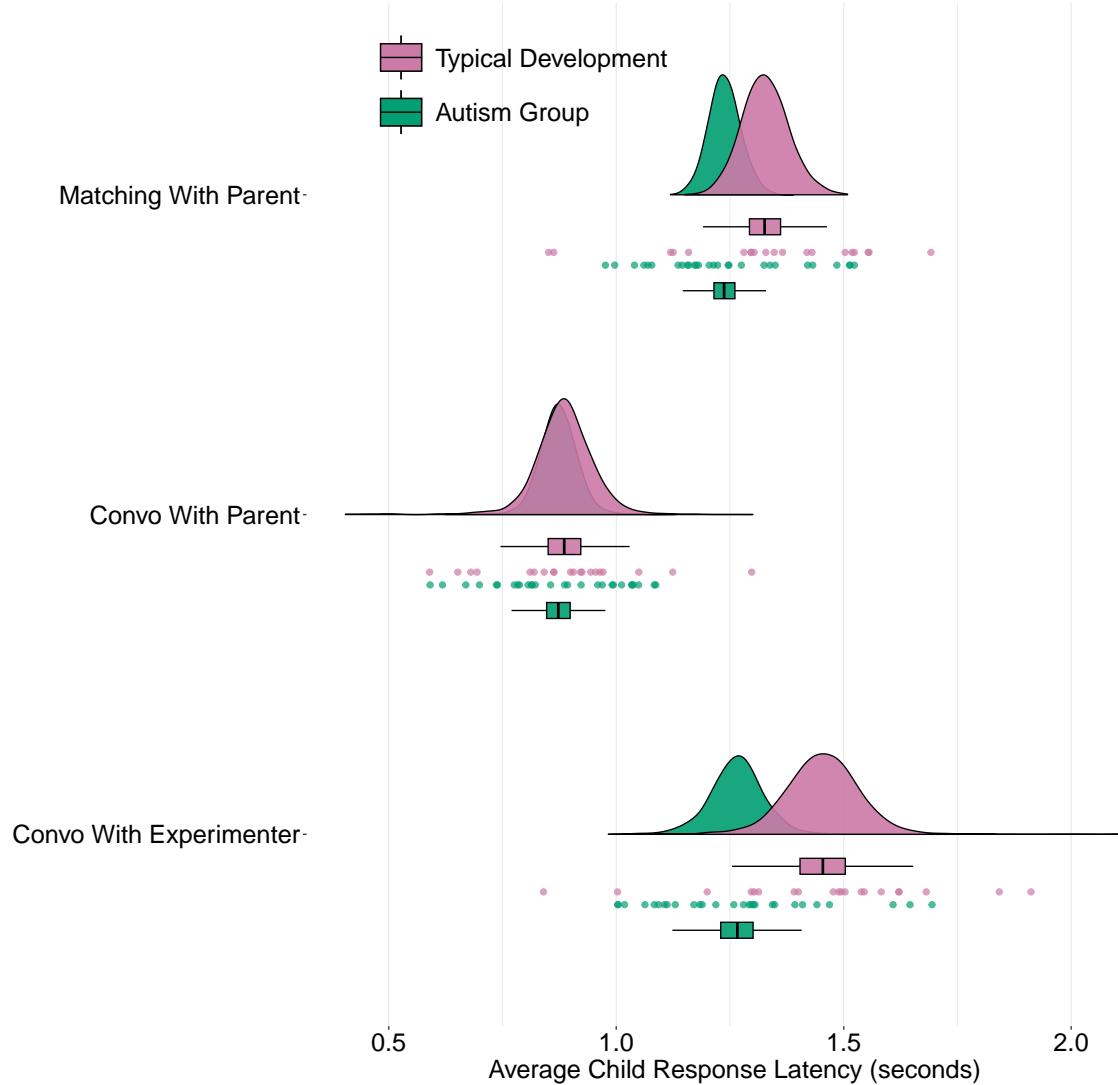


Figure S6: Model estimates for child response latencies across different conversational contexts for autistic children (green) and typically developing children (purple) when excluding overlaps from the model. The points represent participant-level posterior predictions from the model. The density plots and boxplots show aggregated posterior predictions for the three different social contexts: Matching Game, Convo With Parent (Conversations with Parents) and Convo With Experimenter (Conversations with Experimenter). The boxplots show the median (vertical line deviding the box), interquartile range (box edges) and most extreme datapoints (whiskers) for the aggregated posterior predictions.

S.1.3.2 Adult Latency With No Overlaps

Table S6: Posterior estimates for adult response latencies across individual conditions (*Matching Game*, *Convo With Parent* (*Conversations with Parents*) and *Convo With Experimenter* (*Conversations with Experimenter*) and aggregated across conditions (*Aggregate Estimate*) for different model parameters: the Gaussian component (*Latency*) in milliseconds, Standard Deviation (*Sigma*) in milliseconds, and exponential component (*Beta*, or long pauses) on log scale.

	Tasks	Autism Group	Typical Development
Latency	Matching With Parent	1195ms [1135, 1257]	1192ms [1092, 1298]
	Convo With Parent	942ms [853, 1027]	1021ms [898, 1157]
	Convo With Experimenter	1135ms [1032, 1240]	1144ms [1009, 1279]
	Aggregate Estimate	1090ms [1036, 1142]	1119ms [1046, 1196]
Beta	Matching With Parent	0.17 [0.12, 0.22]	0.14 [0.05, 0.22]
	Convo With Parent	-0.09 [-0.16, -0.01]	-0.06 [-0.18, 0.07]
	Convo With Experimenter	-0.34 [-0.47, -0.21]	-0.31 [-0.43, -0.19]
	Aggregate Estimate	-0.08 [-0.14, -0.03]	-0.07 [-0.14, -0.01]
Sigma	Matching With Parent	6ms [3.1, 9.5]	7ms [3.73, 11.57]
	Convo With Parent	9ms [4.62, 14.61]	15ms [9, 26]
	Convo With Experimenter	249ms [217, 284]	234ms [197, 276]
	Aggregate Estimate	23ms [17, 30]	29ms [22, 37]

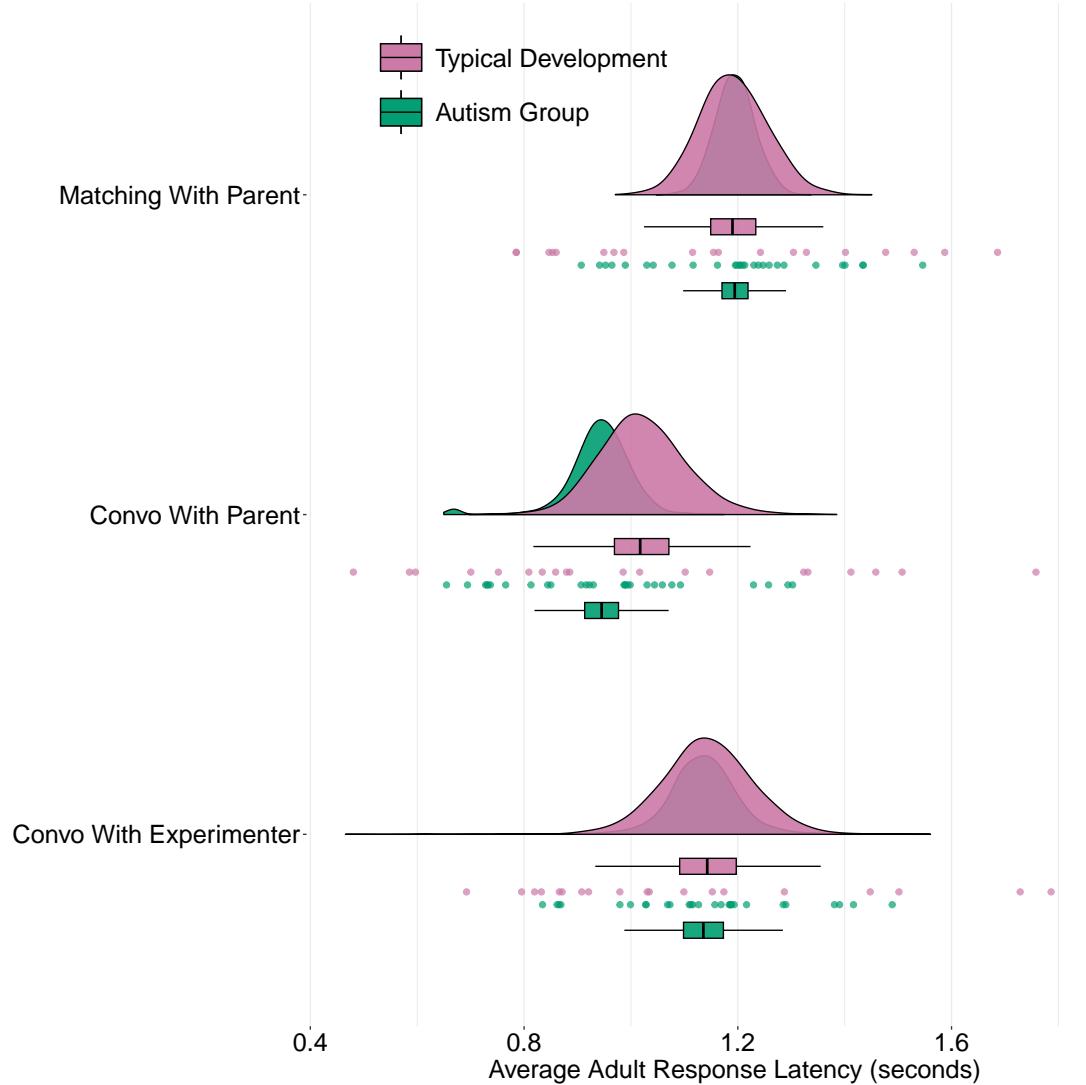


Figure S7: Model estimates for adult response latencies across different conversational contexts for autistic children (green) and typically developing children (purple) when excluding overlaps from the model. The points represent participant-level posterior predictions from the model. The density plots and boxplots show aggregated posterior predictions for the three different social contexts: Matching Game, Convo With Parent (Conversations with Parents) and Convo With Experimenter (Conversations with Experimenter). The boxplots show the median (vertical line deviding the box), interquartile range (box edges) and most extreme datapoints (whiskers) for the aggregated posterior predictions

S.1.3.3 Child Latency Across Gender

Table S7: Posterior estimates for child response latencies across genders and individual conditions (*Matching Game*, *Convo With Parent* (*Conversations with Parents*) and *Convo With Experimenter* (*Conversations with Experimenter*) and aggregated across conditions (*Aggregate Estimate*).

	Male	Female	NonBinary
Autism Group			
Matching With Parent	892ms [755, 1029]	830ms [679, 981]	740ms [428, 1041]
Convo With Parent	621ms [496, 748]	521ms [356, 690]	412ms [70, 773]
Convo With Experimenter	793ms [646, 938]	738ms [552, 926]	696ms [286, 1114]
Aggregate	769ms [672, 864]	697ms [577, 817]	616ms [357, 883]
Typical Development			
Matching With Parent	1033ms [853, 1213]	895ms [748, 1040]	
Convo With Parent	719ms [505, 934]	626ms [454, 799]	
Convo With Experimenter	941ms [670, 1208]	1061ms [829, 1282]	
Aggregate	898ms [743, 1051]	860ms [737, 982]	

S.1.3.4 Child Latency With No Sigma and No Beta

To explore how the estimates change with varying beta and sigma components, we ran a control model without the varying beta and sigma components. The posterior estimates from this model can be viewed in **Table S8** below.

Table S8: *Posterior estimates for child response latencies across individual conditions (Matching Game, Convo With Parent (Conversations with Parents) and Convo With Experimenter (Conversations with Experimenter) and aggregated across conditions (Aggregate Estimate) for different model parameters without a varying beta and sigma. SD refers to standard deviation of intercepts according to child and visit.*

	Tasks	Autism Group	Typical Development
Latency	Matching With Parent	785ms [704, 868]	813ms [729, 899]
	Convo With Parent	726ms [585, 867]	832ms [662, 1002]
	Convo With Experimenter	766ms [629, 909]	970ms [761, 1177]
	Aggregate Estimate	759ms [672, 844]	872ms [764, 981]
Child SD	Matching With Parent	204ms [157, 261]	195ms [137, 269]
	Convo With Parent	210ms [157, 277]	340ms [253, 447]
	Convo With Experimenter	352ms [277, 442]	460ms [354, 589]
	Aggregate Estimate	256ms [215, 303]	332ms [276, 395]
Visit SD	Matching With Parent	69ms [25, 134]	42ms [3.62, 100.02]
	Convo With Parent	115ms [16, 308]	78ms [4.7, 254.14]
	Convo With Experimenter	81ms [8, 214]	114ms [26, 273]
	Aggregate Estimate	88ms [37, 172]	78ms [30, 158]

S.1.3.5 Child Latency With Sigma But No Beta

To explore how the estimates change with a varying beta component, we ran a control model without the varying beta component but retained the varying sigma component. The posterior estimates from this model can be viewed in **Table S9** below.

Table S9: *Posterior estimates for child response latencies across individual conditions (Matching Game, Convo With Parent (Conversations with Parents) and Convo With Experimenter (Conversations with Experimenter) and aggregated across conditions (Aggregate Estimate) for different model parameters with a varying sigma but no beta.*

	Tasks	Autism Group	Typical Development
Latency	Matching With Parent	758ms [679, 839]	786ms [711, 868]
	Convo With Parent	699ms [584, 809]	793ms [642, 947]
	Convo With Experimenter	830ms [699, 957]	1057ms [855, 1256]
	Aggregate Estimate	762ms [683, 837]	879ms [779, 982]
Sigma	Matching With Parent	645ms [599, 694]	569ms [520, 622]
	Convo With Parent	495ms [440, 557]	571ms [496, 660]
	Convo With Experimenter	897ms [834, 963]	879ms [808, 956]
	Aggregate Estimate	659ms [626, 695]	659ms [617, 701]
Child SD	Matching With Parent	189ms [145, 242]	183ms [130, 251]
	Convo With Parent	204ms [151, 267]	308ms [233, 404]
	Convo With Experimenter	328ms [259, 410]	453ms [340, 592]
	Aggregate Estimate	240ms [202, 282]	315ms [260, 377]
Visit SD	Matching With Parent	72ms [31, 134]	38ms [2.98, 93.19]
	Convo With Parent	82ms [6, 248]	80ms [4.95, 248.07]
	Convo With Experimenter	84ms [11, 213]	94ms [9, 240]
	Aggregate Estimate	79ms [34, 153]	71ms [23, 147]

S.1.3.6 Child Overlaps according to Individual Differences

To explore the extent to which potential increases or decreases in response latencies were a function of a greater proportion of overlaps, we modelled the proportion of overlaps (i.e., negative latencies) according to individual differences among children, as shown in **Figure S8**.

The Bayesian logistic regression with a Bernoulli likelihood revealed several credible individual difference effects on overlap probability. Most notably, cognitive skills showed a strong positive association with overlaps for typically developing children in conversations with their parents ($\beta = 1.13 [0.08, 2.14]$, ER = 52). A smaller but credible positive effect of cognitive skills was also observed for children with ASD during familiar question tasks ($\beta = 0.42 [0.00, 0.85]$, ER = 40). These findings suggest that children with better cognitive abilities were substantially more likely to produce overlapping speech during familiar question contexts.

In contrast, motivation showed diverging effects between diagnostic groups. For children with autism, higher motivation was associated with reduced overlap probability in conversations with parents ($\beta = -0.27 [-0.57, 0.03]$, ER = 14) and the matching game ($\beta = -0.26 [-0.51, -0.01]$, ER = 22), while the effects for typically developing children were positive but less certain (e.g., for matching game: $\beta = 0.31 [-0.21, 0.84]$, ER = 5). This pattern suggests that motivation may lead to more controlled responding in children with autism but potentially faster, more overlapping responses in typically developing children.

Additionally, awareness skills showed a credible negative effect for typically developing children during unfamiliar question tasks ($\beta = -0.70 [-1.53, 0.16]$, ER = 18), suggesting that higher awareness may lead to more controlled responding that reduces overlaps in challenging contexts. Other individual difference measures, including language, and motor skills, showed weaker or non-credible associations with overlap probability, with most 95% credible intervals including zero. Overall, cognitive skills emerged as the strongest predictor of overlap behaviour, particularly for typically developing children in familiar conversational

contexts.

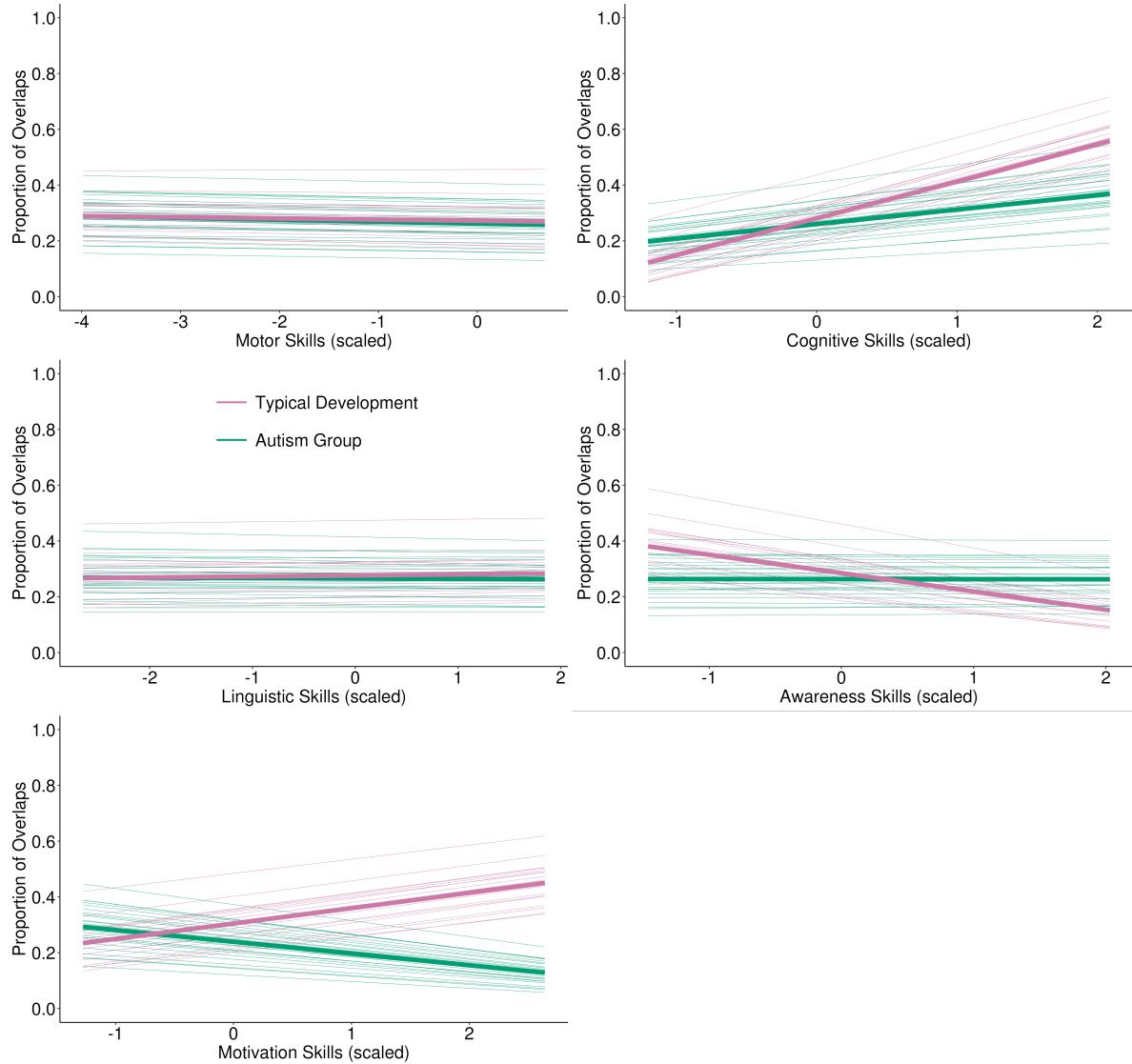


Figure S8: This panel of plots displays the proportion of overlaps according to individual differences of the children. The faded lines are posterior predictions from model for individual subjects and how the proportion of overlaps changes for each of the continuous predictors.

S.1.3.7 Child Individual Differences With No Overlaps

To check the extent to which overlaps were driving the estimates of change in response latencies according to individual differences, we ran the same individual differences model on data where overlaps (i.e., negative latencies) were excluded. The posterior estimates from this model can be viewed in **Table S10** below.

Table S10: *Posterior estimates for models when excluding overlaps for how child response latencies change across individual conditions (Matching Game, Parent Conversations and Experimenter Conversations) and all conditions (Aggregate Estimate) as a function of each of the types of skills listed below (Motor, Cognitive Skills, Language Skills, Social Awareness and Social Motivation). The estimates denote the change in child response latency as a function of one standard deviation increase in the skills of the children.*

	Skills	Autism Group	Typical Development
Social Cognition	Matching With Parent	-13ms [-42, 15]	-95ms [-382, 147]
	Convo With Parent	12ms [-30, 55]	50ms [-171, 291]
	Convo With Experimenter	-76ms [-176, 27]	-490ms [-883, -83]
	Aggregate Estimate	-26ms [-67, 16]	-178ms [-367, -4]
Social Awareness	Matching With Parent	4.24ms [-29, 37]	69ms [-81, 252]
	Convo With Parent	27ms [-21, 77]	-59ms [-205, 81]
	Convo With Experimenter	-4.11ms [-117, 108]	190ms [-65, 449]
	Aggregate Estimate	9ms [-39, 56]	67ms [-44, 189]
Social Motivation	Matching With Parent	4.48ms [-21, 30]	-46ms [-219, 116]
	Convo With Parent	-26ms [-64, 11]	78ms [-77, 245]
	Convo With Experimenter	84ms [-3.38, 169.97]	90ms [-194, 377]
	Aggregate Estimate	21ms [-15, 57]	40ms [-84, 165]
Language	Matching With Parent	1.94ms [-14, 18]	20ms [-33, 84]
	Convo With Parent	16ms [-8, 43]	-18ms [-73, 35]
	Convo With Experimenter	0.01ms [-57, 58]	35ms [-59, 131]
	Aggregate Estimate	6ms [-18, 31]	12ms [-29, 55]
Motor	Matching With Parent	0.18ms [-18, 18]	11ms [-69, 94]
	Convo With Parent	-6ms [-35, 21]	-9ms [-89, 67]
	Convo With Experimenter	41ms [-21, 101]	-15ms [-155, 130]
	Aggregate Estimate	12ms [-14, 36]	-4.13ms [-67, 57]

S.1.3.8 Predictability

As shown in **Table S11**, the overall degree of utterance predictability differed slightly across interactional contexts, with both the child and adult utterances agreeing with each other in terms of their overall degree of predictability.

Table S11: *Overview of Average Utterance Predictability Across Different Contexts*

Condition	Diagnosis	Adult Predictability	Child Predictability
Matching With Parent	Autism Group	0 (0.98)	-0.02 (0.96)
Matching With Parent	Typical Development	-0.02 (1.02)	-0.07 (0.97)
Convo With Parent	Autism Group	0.08 (1.05)	0.19 (1.02)
Convo With Parent	Typical Development	0.08 (1.1)	0.23 (1.04)
Convo With Experimenter	Autism Group	-0.08 (1)	0.01 (0.98)
Convo With Experimenter	Typical Development	-0.15 (0.97)	-0.03 (0.98)

S.1.3.9 Child Latency With Predictability, No Short Utterances

To check the potential influence of shorter utterances on the posterior estimates, we removed all utterances under three words in the corpus and ran the same predictability model as in the main manuscript. The estimates from this control model are shown in **Table S12**.

Table S12: *Posterior estimates for how child response latencies change across individual conditions (Matching With Parent, Parent Conversations and Experimenter Conversations) and all conditions (Aggregate Estimate) as a function of the predictability of the previous adult utterance and child utterance. The estimates denote the change in child response latency as a function of one standard deviation increase in the predictability of the utterance (i.e., cosine similarity).*

	Tasks	Autism Group	Typical Development
Predictability of Previous Adult Utterance	Matching With Parent	-54ms [-93, -14]	33ms [-18, 86]
	Convo With Parent	-25ms [-66, 17]	-72ms [-136, -6]
	Convo With Experimenter	-36ms [-93, 21]	-11ms [-83, 61]
	Aggregate Estimate	-38ms [-66, -11]	-17ms [-53, 20]
Predictability of Own Utterance	Matching With Parent	-31ms [-60, -3]	-7ms [-47, 32]
	Convo With Parent	10ms [-26, 47]	-12ms [-58, 34]
	Convo With Experimenter	188ms [145, 230]	108ms [50, 166]
	Aggregate Estimate	56ms [34, 77]	29ms [1.13, 56.88]

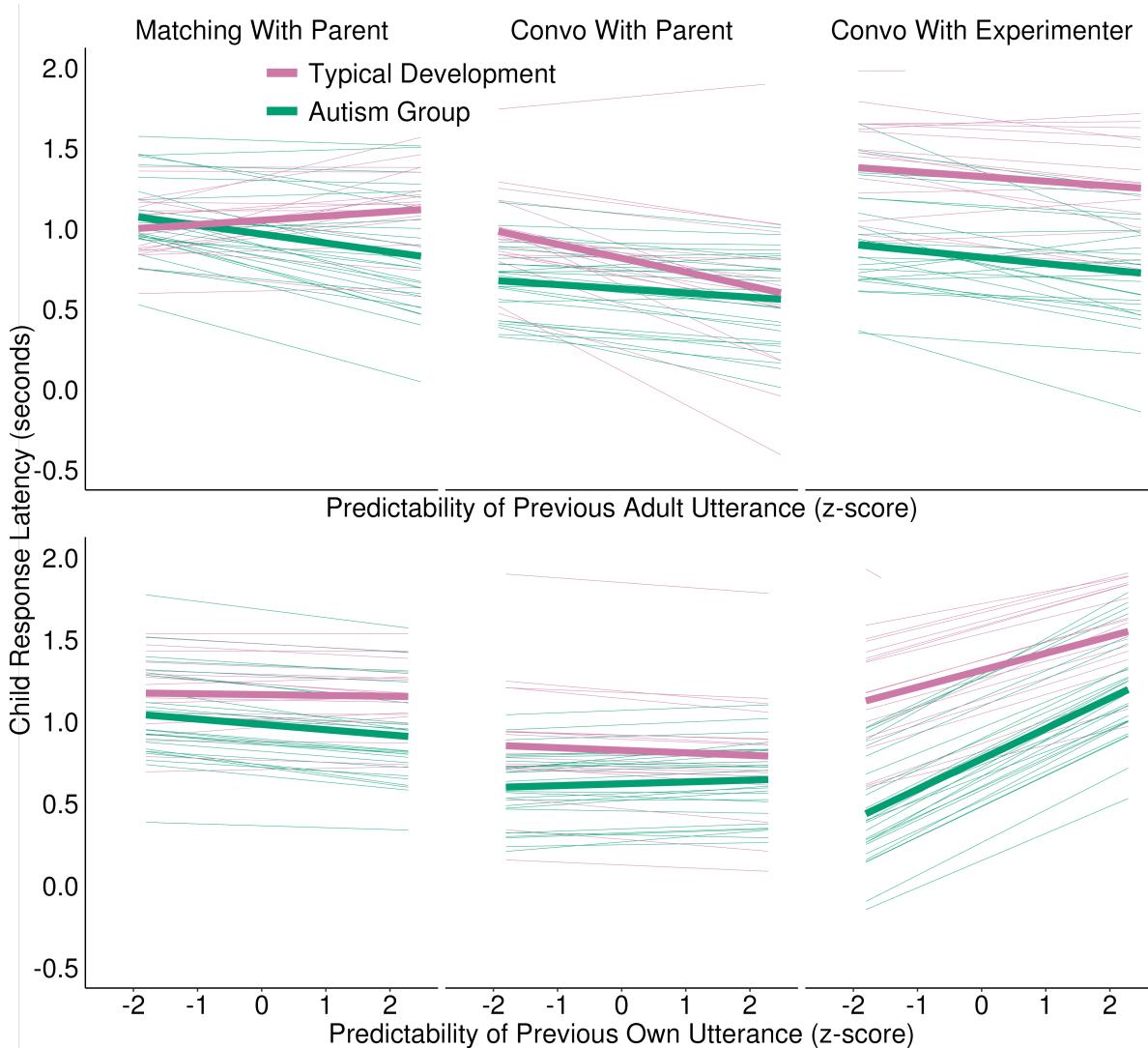


Figure S9: Posterior estimates in a model excluding utterances below three words. The plot shows how child response latencies change across individual conditions (Matching Game, Parent Conversations and Experimenter Conversations) and all conditions (Aggregate Estimate) as a function of the predictability of the previous adult utterance and current child utterance.

S.1.3.10 Child Latency With Increasing Familiarity

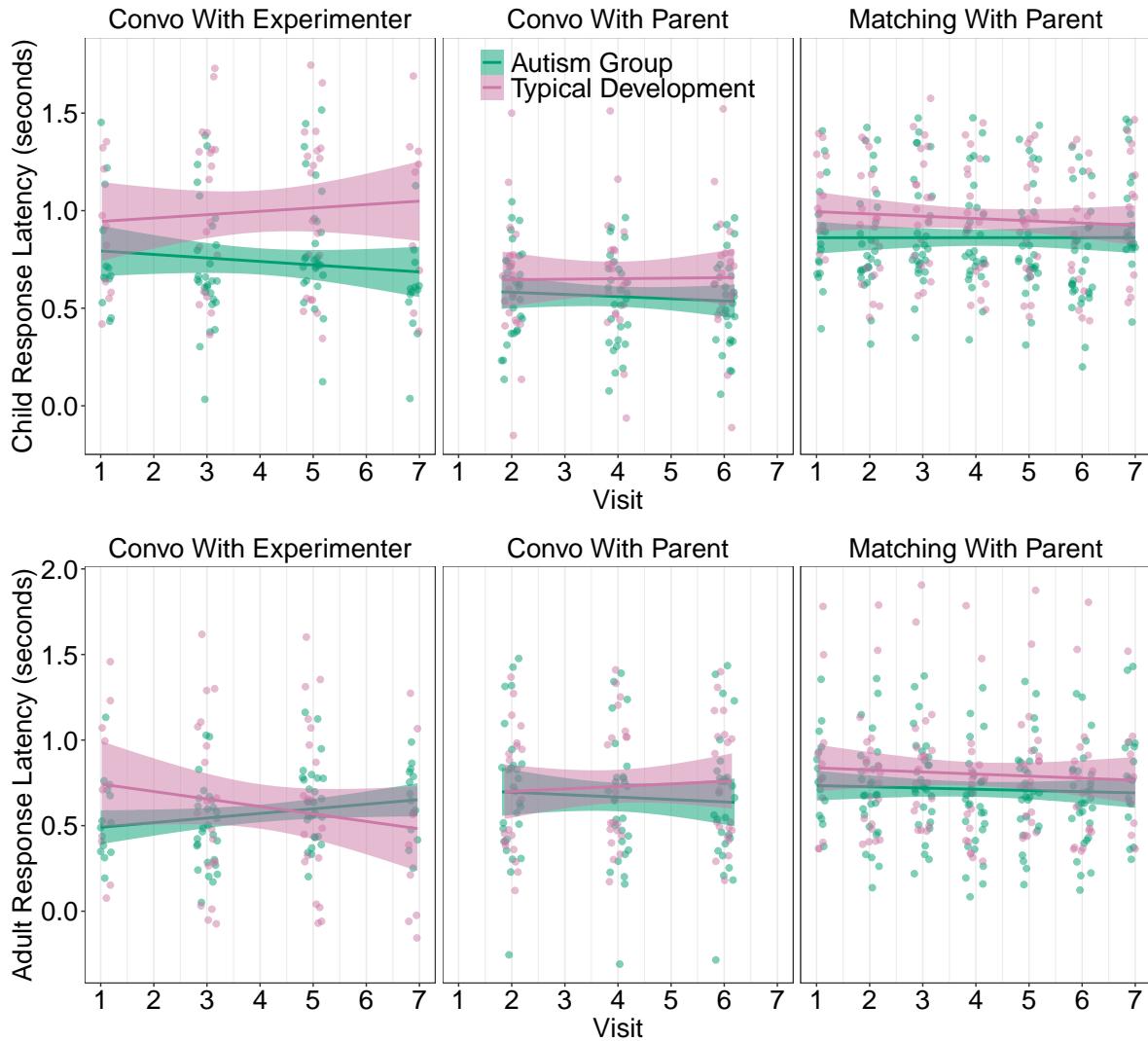


Figure S10: Posterior estimates for child response latencies across individual conditions (Matching Game, Parent Conversations and Experimenter Conversations) as a function of visit. Individual points denote participant-level posterior estimates from the model for the children (top) and adults (bottom). Note that conversations took place alternately with caregivers (in even weeks) and the experimenter (in odd weeks).

S.1.3.11 Child Latency For Surrogate Pairs

As a control analysis, we ran the first model after rotating the dataset; that is, pairing the utterances of one child with the interactant of another child and recalculating the response latencies between turns. The estimates from this control model are shown in **Table S13**.

Table S13: *Posterior estimates for child response latencies across individual conditions (Matching Game, Convo With Parent and Experimenter Conversations) and aggregated across conditions (Aggregate Estimate) for different model parameters: the Gaussian component (Latency), exponential component (Beta), Residual Heterogeneity (Sigma), between-subjects standard deviation (Child SD), standard deviation across visits (Visit SD), Residual Heterogeneity across visits (Sigma Visit) and proportion of latencies below zero (Overlap Proportion).*

	Tasks	Autism Group	Typical Development
Latency	Matching With Parent	27ms [-1632, 1806]	-202ms [-1896, 1544]
	Convo With Parent	-1823ms [-2576, -1087]	-2070ms [-3003, -1079]
	Convo With Experimenter	-2927ms [-3751, -2156]	-2009ms [-2671, -1330]
	Aggregate Estimate	-1574ms [-2293, -831]	-1427ms [-2183, -632]
Sigma	Matching With Parent	17735ms [13366, 23464]	16909ms [12333, 23217]
	Convo With Parent	4737ms [3299, 6966]	4662ms [2355, 8587]
	Convo With Experimenter	8091ms [5725, 11433]	5827ms [3685, 9164]
	Aggregate Estimate	8793ms [7070, 10988]	7716ms [5553, 10557]
Child SD	Matching With Parent	3015ms [1756, 4627]	3100ms [1917, 4587]
	Convo With Parent	1102ms [744, 1526]	823ms [458, 1261]
	Convo With Experimenter	1579ms [1070, 2196]	1096ms [687, 1611]
	Aggregate Estimate	1899ms [1405, 2496]	1673ms [1224, 2220]

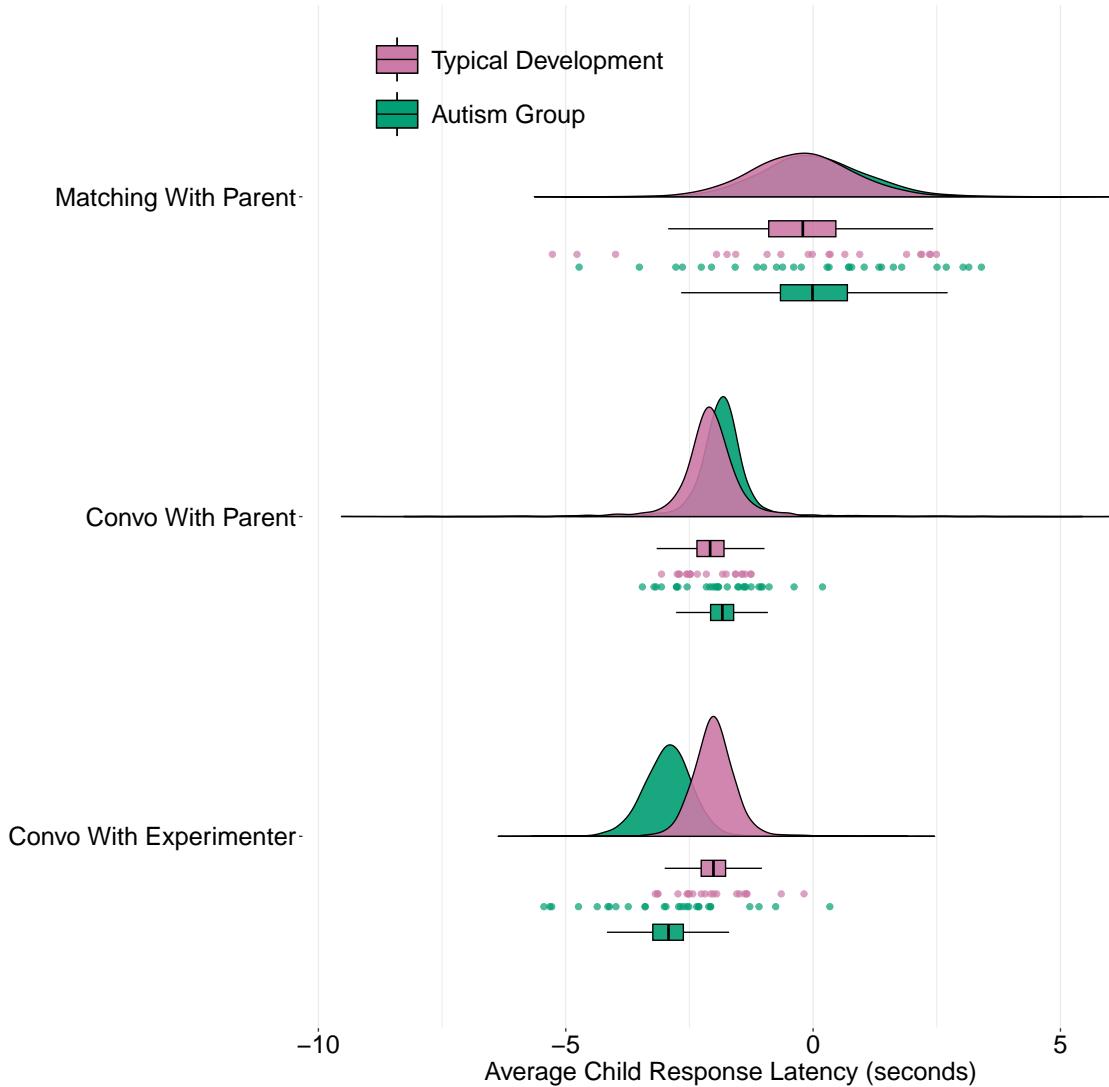


Figure S11: Posterior estimates for surrogate child response latencies across individual conditions (Matching Game, Parent Conversations and Experimenter Conversations).

S.1.4. Model Information and Quality Checks

S.1.4.1 Choice of Priors

We chose weakly informative priors in order to ensure that their influence on the estimates was small and to discount extreme effect sizes as unlikely (Lemoine, 2019; Gelman et al., 2017). For the overall distribution, we chose an ex-Gaussian distribution. Our prior for the Gaussian component was specified as having a mean of 1 and standard deviation of 1 based on our prior expectations for child response latencies in a turn-taking context (Nguyen et al., 2022). This prior implies that we expect approximately 95% of the child latencies to fall between -1sec and 3sec. For the between-subject variability across different conversational contexts, we encoded our expectations with a truncated (i.e., standard deviation must have positive values) Gaussian prior with a mean of 0 and a standard deviation of 0.3, which implies that we expect the vast majority of values for the between-subject variability within diagnostic group to be between 0 and 0.6sec. For the rate parameter of the exponential component of the distribution, we specified a Gaussian prior with a mean of 0 and standard deviation of 1 and allowed this to vary for each task and familiarity for each participant.

The models were fitted with Hamiltonian Monte Carlo samplers with 2 parallel chains with 5,000 iterations each, an adapt delta of 0.99 and a maximum tree depth of 20 in order to ensure no divergence in the estimation process. The quality of the models was assessed by (i) ensuring Rhat statistics to be lower than 1.1, (ii) carrying out prior and posterior predictive checks, (iii) plotting prior against posterior estimates and assessing whether the posteriors had lower variance than the priors, (iv) ensuring no divergences in the process of estimation, (v) checking that the number of effective bulk and tail samples was above 200, (vi) conducting prior sensitivity analyses.

All models were fitted using the R package *brms* (Bürkner, 2017). As an example of what the basic *brms* formula looked like, here is the formula for the first model (cf. code on OSF for the full code and *brms* formulae for other models). Note that visit was excluded

from fixed effects for the location parameter as temporal changes in mean response were adequately modeled through random slope structures, while visit-specific heteroscedasticity was explicitly parameterized through fixed effects in the scale parameter (i.e., sigma):

$$\text{Latency} \sim 0 + \text{Diagnosis:Task:Familiarity} + (0 + \text{Task:Familiarity} \mid p \mid \text{gr}(ID, by = Diagnosis)) + (0 + \text{Diagnosis:Task:Familiarity} \mid r \mid \text{Visit}),$$

$$\text{Sigma} \sim 0 + \text{Diagnosis:Task:Familiarity} + \text{Diagnosis:Task:Familiarity:Visit} + (0 + \text{Task:Familiarity:Visit} \mid p \mid \text{gr}(ID, by = Diagnosis)),$$

$$\text{Beta} \sim 0 + \text{Diagnosis:Task:Familiarity} + (0 + \text{Task:Familiarity} \mid p \mid \text{gr}(ID, by = Diagnosis)) + (0 + \text{Diagnosis:Task:Familiarity} \mid r \mid \text{Visit})$$

S.1.4.2 Prior and Posterior Predictive Checks

We performed quality checks of the models by carrying out prior and posterior predictive checks. The below prior predictive checks (on the left) indicate that our priors predict values within the order of magnitude of the distribution. The posterior predictive checks (on the right) indicate that the models have captured the distributions of data. These plots provide reassurance that our models capture relevant aspects of the overall distributions of dependent variables.

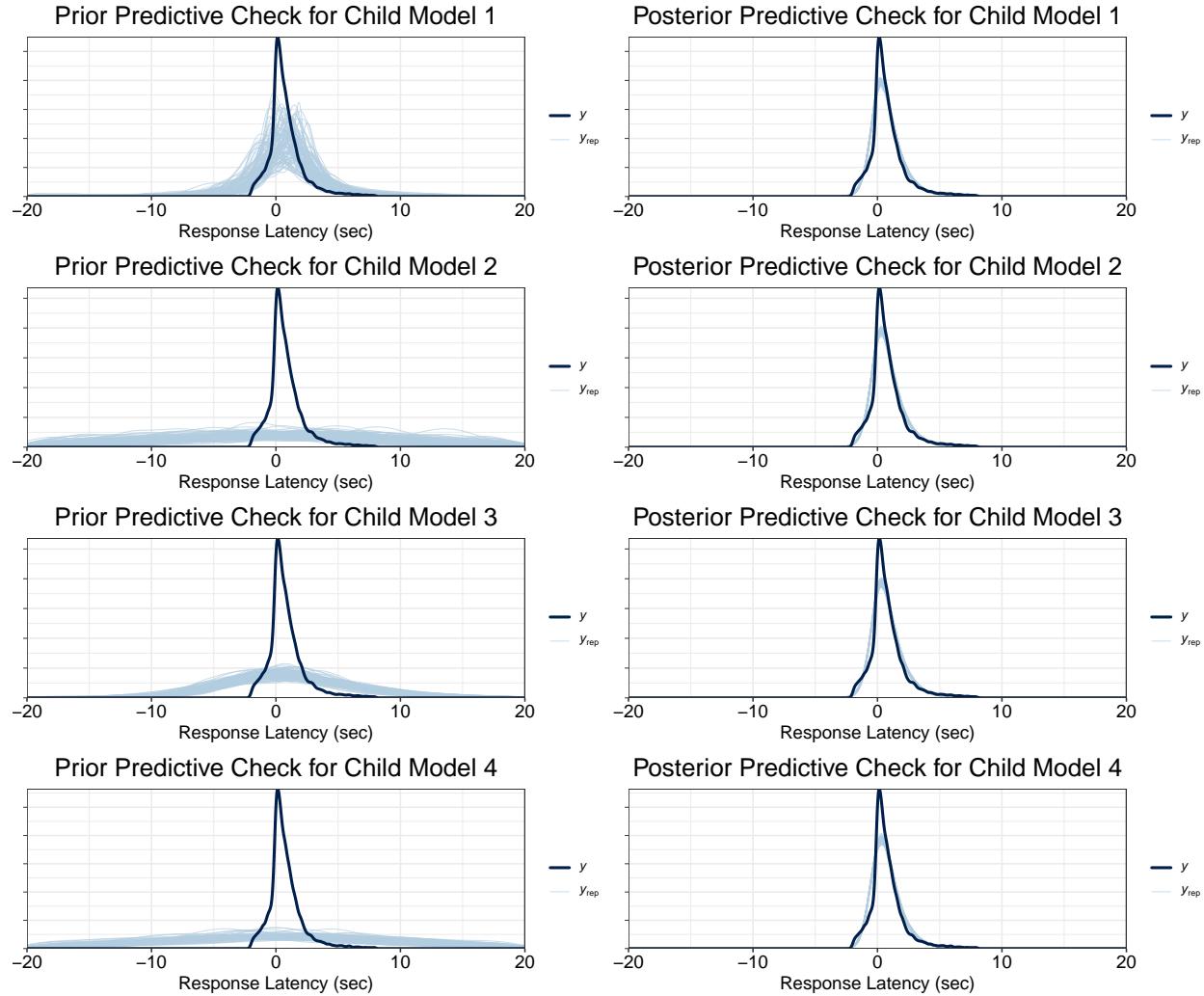


Figure S12: Prior predictive checks (left column) and posterior predictive checks (right column) for child latency in models 1-4.

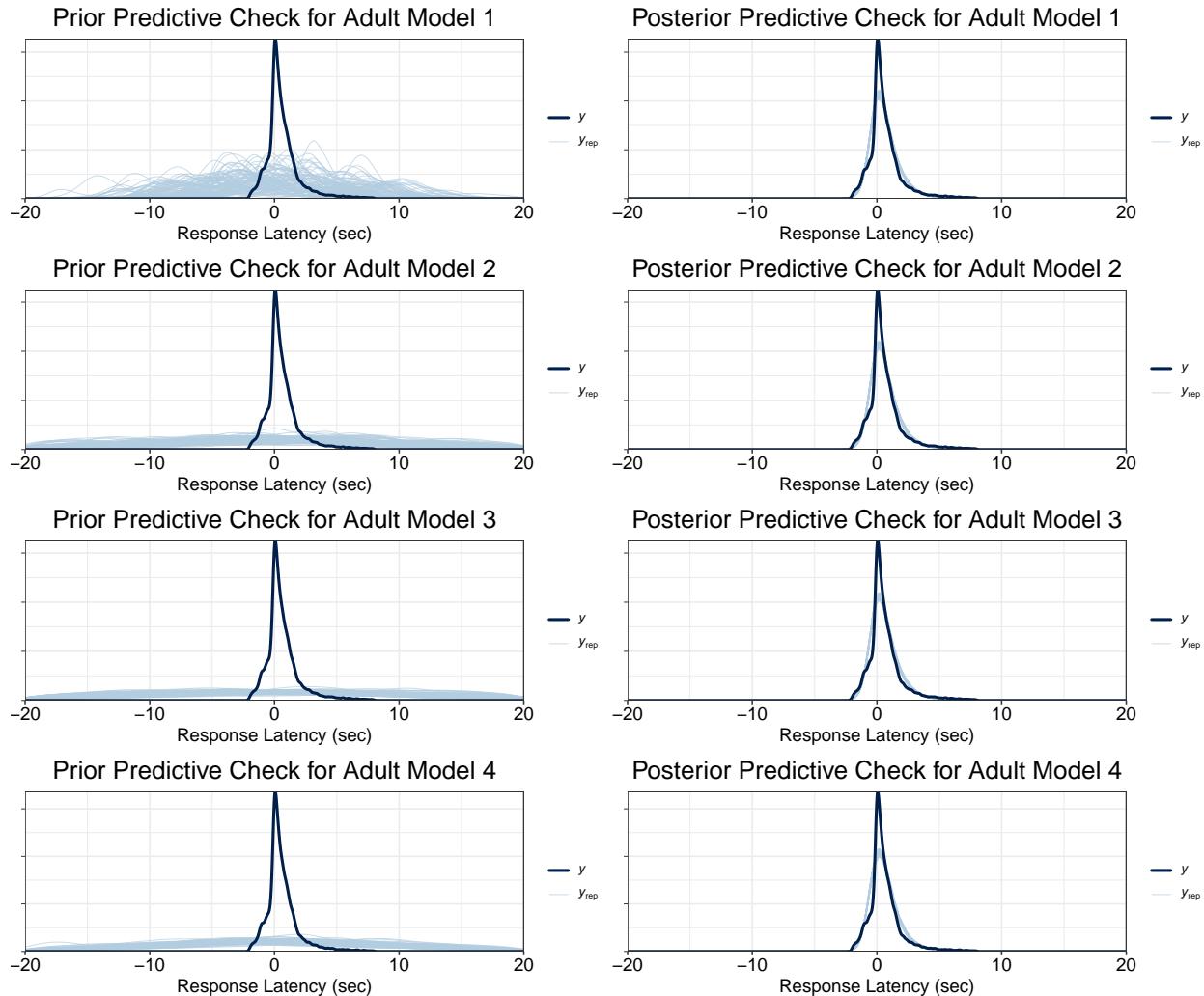


Figure S13: Prior predictive checks (left column) and posterior predictive checks (right column) for adult latency in models 1-4.

S.1.4.3 Prior-Posterior Update Plots

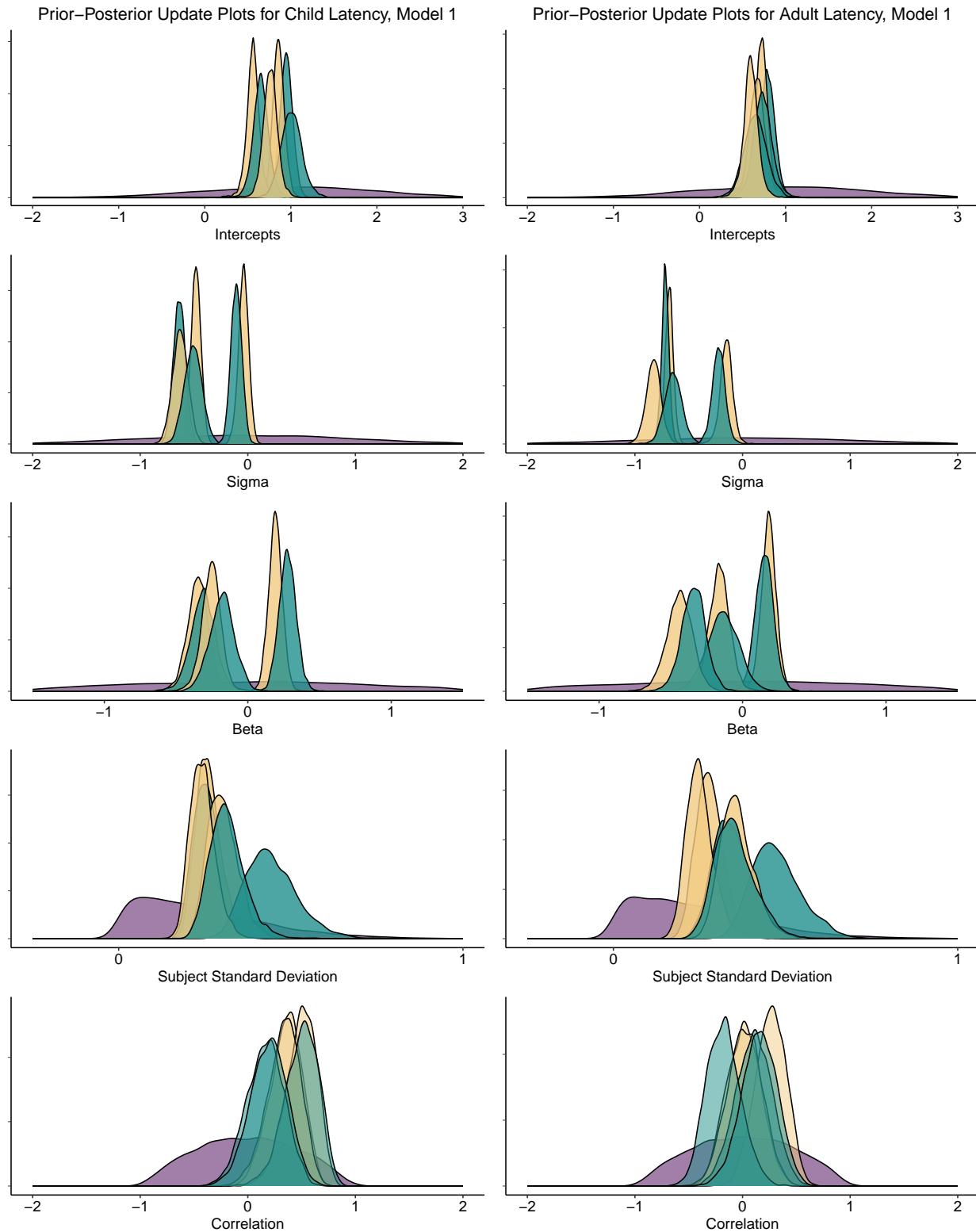


Figure S14: Prior-posterior update checks for child and adult latency in Model 1. Purple density plots show the prior predictive density plot, turquoise density plots indicate posterior predictions for the typical development group, and yellow density plots denote the predicted estimates for the autism group.

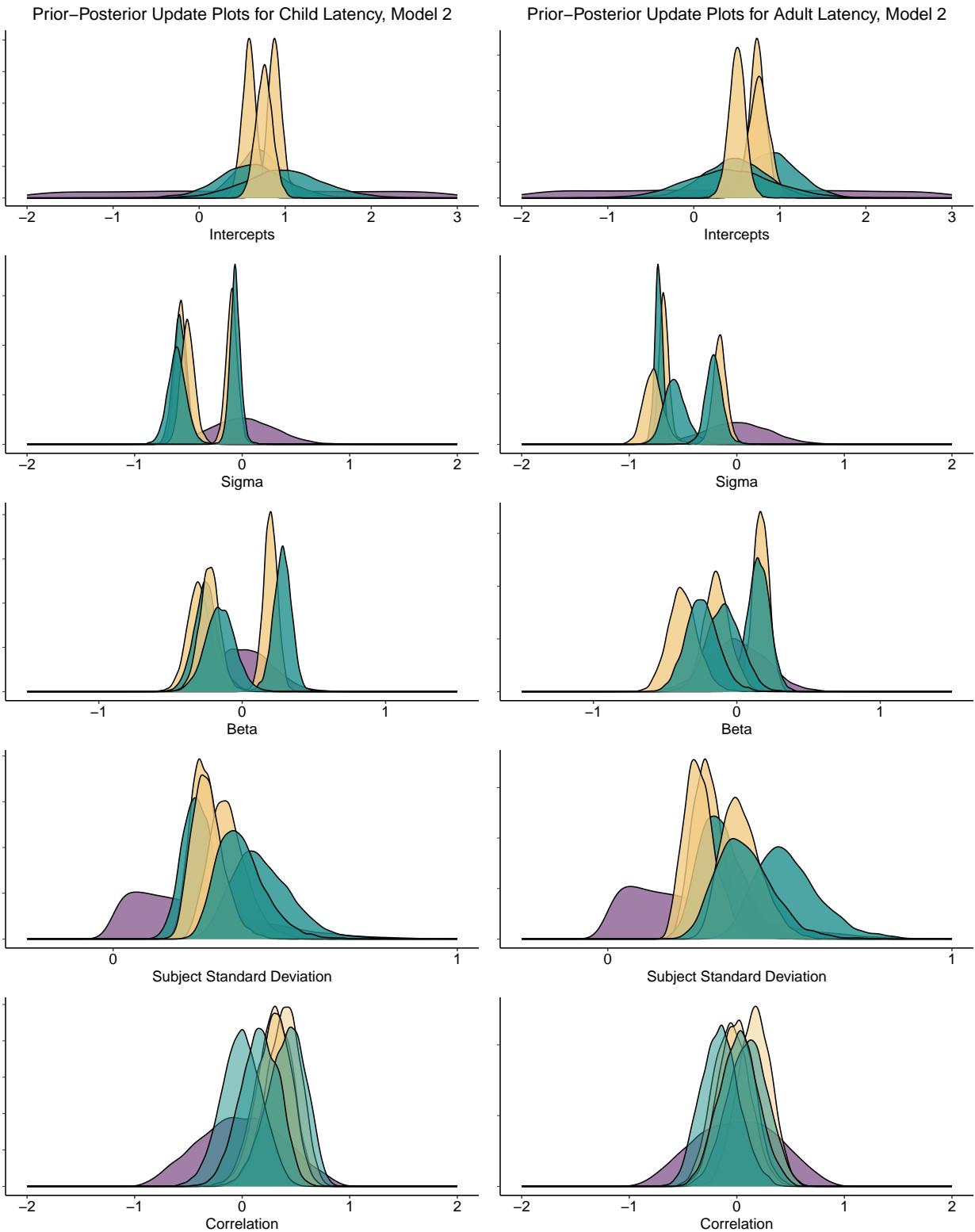


Figure S15: Prior-posterior update checks for child and adult latency in Model 2. Purple density plots show the prior predictive density plot, turquoise density plots indicate posterior predictions for the typical development group, and yellow density plots denote the predicted estimates for the autism group.

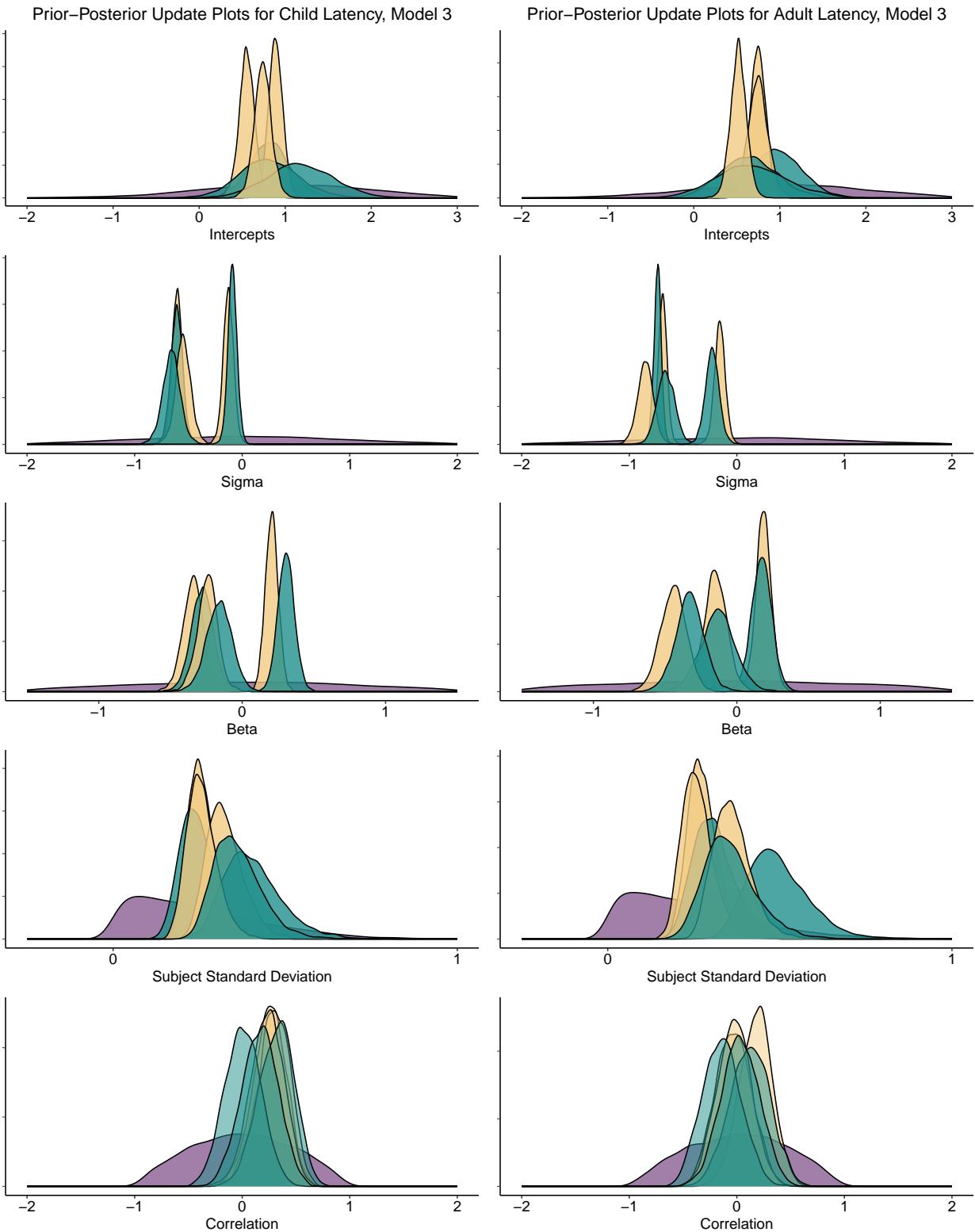


Figure S16: Prior-posterior update checks for child and adult latency in Model 3. Purple density plots show the prior predictive density plot, turquoise density plots indicate posterior predictions for the typical development group, and yellow density plots denote the predicted estimates for the autism group.

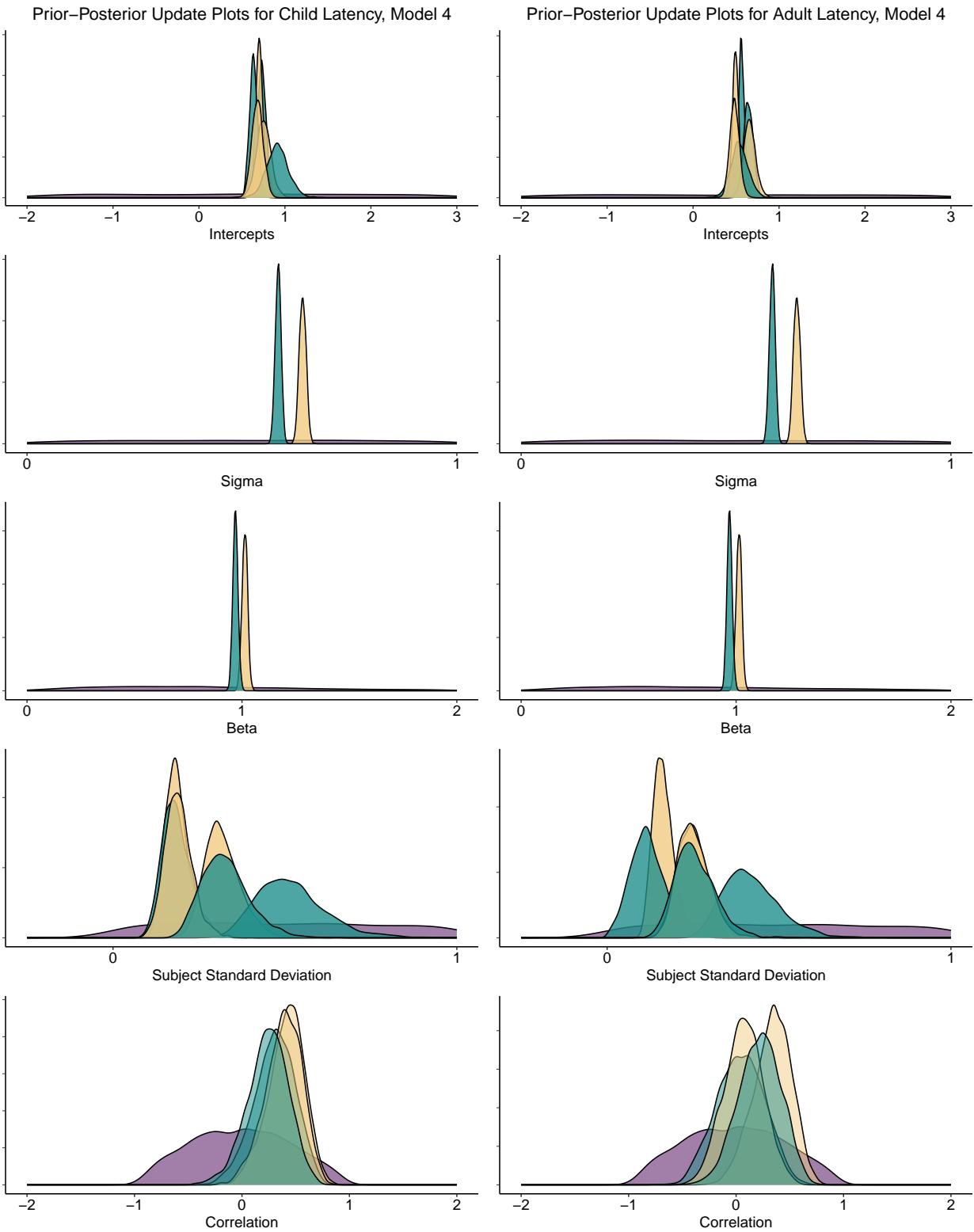


Figure S17: Prior-posterior update checks for child and adult latency in the multivariate Model 4. Purple density plots show the prior predictive density plot, turquoise density plots indicate posterior predictions for the typical development group, and yellow density plots denote the predicted estimates for the autism group.

S.1.5. Time Code Reliability

To assess the reliability of our transcription process, we compared two parallel transcriptions of the same audio recordings. Child speech segments were independently transcribed by two different transcribers (Transcriber 1 and Transcriber 2) across 104 conversation segments from our dataset, representing both structured task contexts (Matching Game) and open-ended conversations. Using a time-based comparison method with 50-millisecond intervals, we calculated Cohen’s Kappa to measure agreement on speech presence/absence across the timeline of each recording, providing a robust measure of inter-transcriber reliability that accounts for chance agreement.

Our analysis revealed excellent agreement between transcribers in identifying child speech segments. The overall mean Cohen’s Kappa was 0.88 (median = 0.89), indicating very high reliability between transcription methods. This level of agreement was consistent across both conversational contexts (mean $\kappa = 0.88$, 95% CI [0.74, 0.97]) and structured Matching Game tasks (mean $\kappa = 0.88$, 95% CI [0.76, 0.97]). The mean percentage agreement across all recordings was 94.7%, demonstrating that transcribers identified nearly identical patterns of child speech throughout the recordings.

Visual inspection of the alignment between transcription methods (**Figure S18**) confirms the strong temporal correspondence between transcribers. The histogram of Kappa values (**Figure S19**) shows that the majority of recordings achieved Kappa values above 0.85, with very few falling below 0.75, indicating consistently high reliability across the dataset. These results demonstrate that our transcription methodology produces highly reliable identification of child speech segments, providing a solid foundation for subsequent analyses of child language and communication patterns in our study.

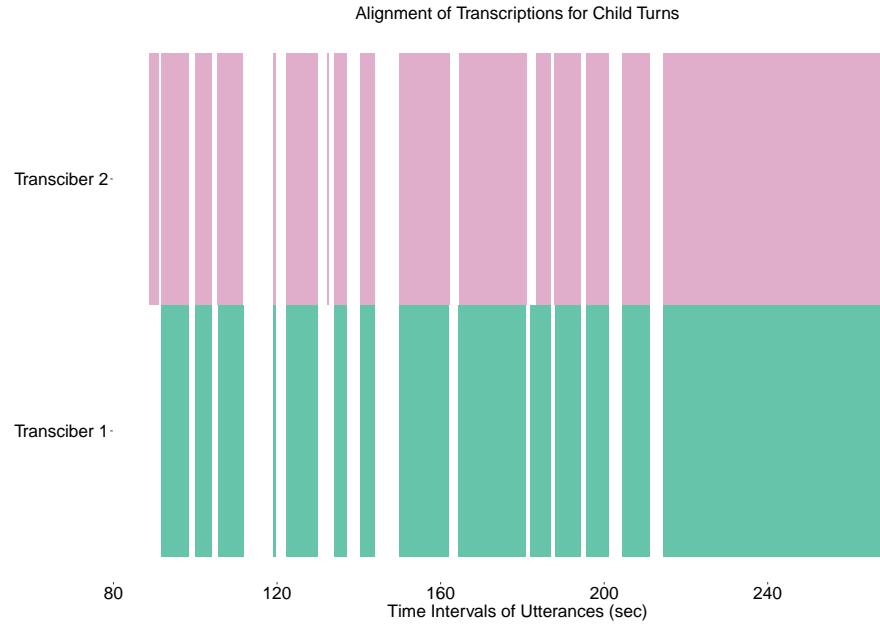


Figure S18: Example of aligned transcriptions for child turns from a representative recording. The temporal alignment between Transcriber 1 (bottom, teal) and Transcriber 2 (top, pink) demonstrates the high degree of agreement in identifying child speech segments.

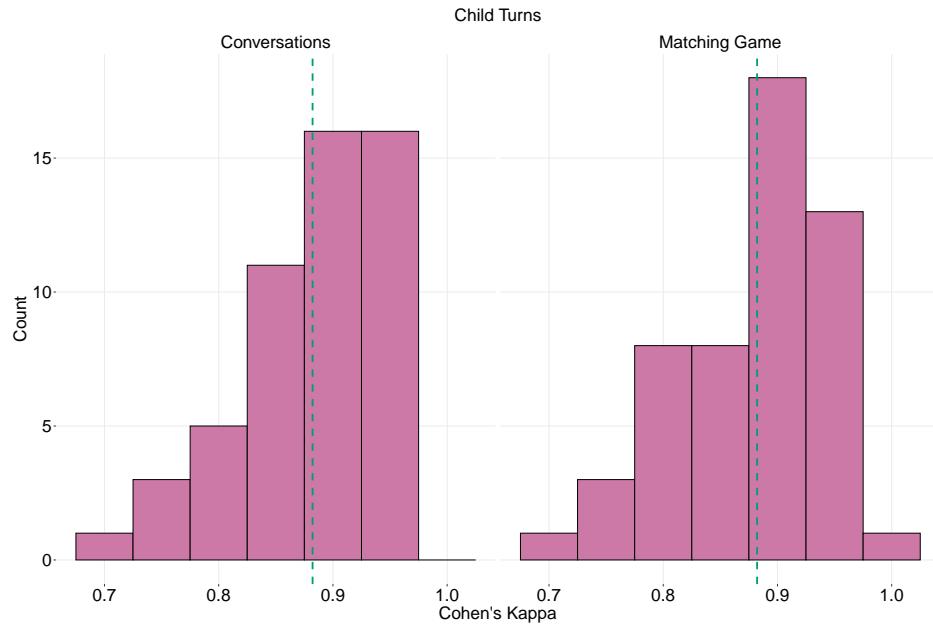


Figure S19: Distribution of Cohen's Kappa values for child turns across different contexts. Left panel shows Kappa values for open-ended conversations ($n=52$); right panel shows Kappa values for structured Matching Game tasks ($n=52$). Vertical dashed lines indicate mean Kappa values for each context.

S.1.6. Recurrence Analysis

Cross-Recurrence Quantification Analysis (CRQA) was employed to quantify non-linear dynamical coupling patterns in dyadic interactions across the three communicative tasks (Matching With Parent, Conversation With Parent, and Conversation With Experimenter) for both the Autism Group and Typical Development Group. CRQA offers a complementary perspective on bidirectional influences on behavioral patterns between interaction partners (R. Fusaroli, Konvalinka, & Wallot, 2014). For the CRQA computation, we used the following parameters: embedding dimension ($m = 1$), delay ($\tau = 1$), and radius ($\epsilon = 0.1 \times$ standard deviation of the normalized time series). These parameters were selected to capture the turn-taking dynamics in our interaction data while ensuring sufficient recurrence points for reliable analysis, and were applied consistently across all interaction samples to ensure comparability.

Figure S20 illustrates the distribution of Recurrence Rate (RR) across the three interaction contexts, demonstrating task-specific patterns of behavioral coordination. A Recurrence Rate (RR) of approximately 35% indicates that a third of all possible points in the interaction time series recur or synchronize, suggesting strong coordination between interaction partners. Table S14 presents a comprehensive set of CRQA metrics that characterize the temporal organization, predictability, and stability of interpersonal coordination in our sample. We report these exploratory analyses here to spur on further investigation in future research specifically designed to test dynamical hypotheses. We also used CRQA to examine directional coupling between child and adult behaviors across multiple time lags (-5 to +5) (see **Figure S21**). In conversations with experimenters, the autism group showed an increasing trend toward positive lags (suggesting adult-led dynamics), while the typical development group appears to exhibit a decline toward positive lags (potentially suggesting less adult-led structure in this context). The parent conversation condition presents a somewhat different picture, where the typical development group appears to demonstrate a possible trend toward adult-led dynamics,

while the autism group shows relatively stable patterns. In the structured matching task with parents, both groups show relatively modest variations across lags. These patterns suggest that interaction dynamics might involve different leadership patterns emerging depending on the social partner and task structure. Further research would need to explore the role of individual differences (e.g. in linguistic and socio-cognitive skills) for these dynamics and how they might impact the experience of the interaction.

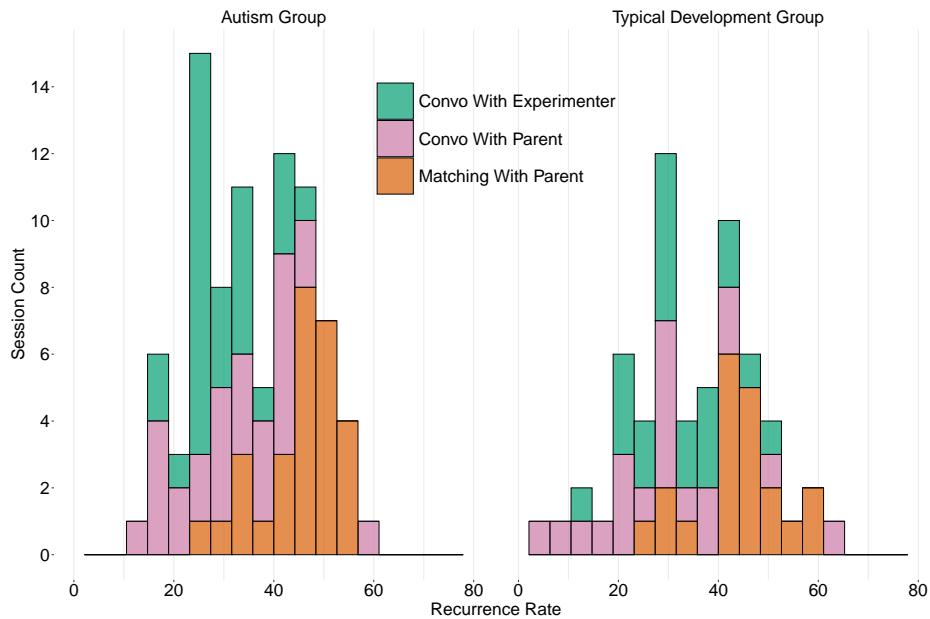


Figure S20: *Distribution of Recurrence Rate (RR) values across interaction tasks by diagnostic group. The histogram shows the frequency distribution of Recurrence Rate values for the Autism Group (left panel) and Typical Development Group (right panel), color-coded by interaction task: Conversation With Experimenter (green), Conversation With Parent (pink), and Matching With Parent (orange).*

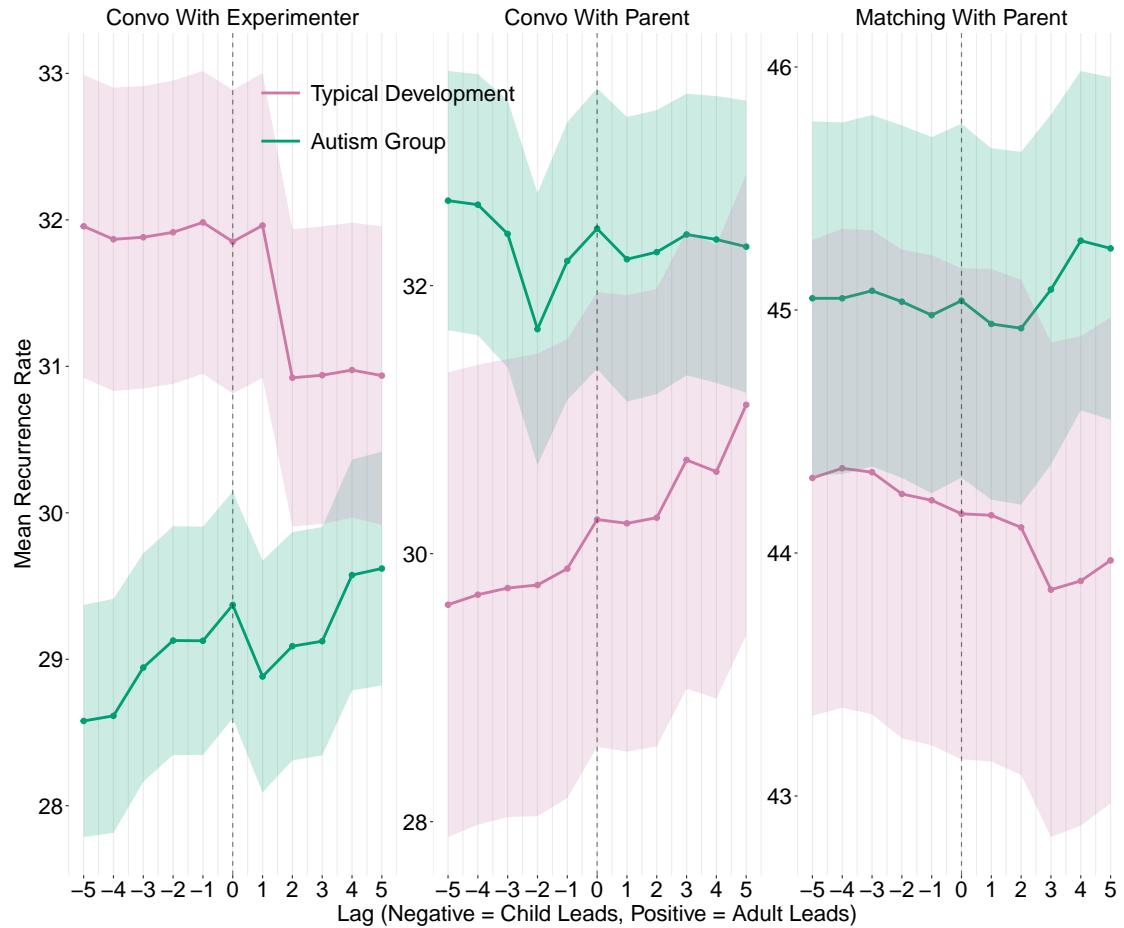


Figure S21: Diagonal recurrence profiles revealing leader-follower dynamics across interaction contexts. Cross-recurrence quantification analysis across time lags (-5 to +5) examining directional coupling in dyadic interactions. Negative lags indicate child-led interactions, positive lags indicate adult-led interactions, and lag 0 represents synchronous patterns. Lines show group means for autism (green) and typical development (pink) groups. Shaded ribbons indicate ± 0.5 standard errors.

Table S14: CRQA metrics quantify different aspects of recurrence patterns in dyadic interactions: *RR* (Recurrence Rate) represents the percentage of recurrent points in the recurrence plot, indicating overall similarity between behavioral patterns of interaction partners; *DET* (Determinism) measures the percentage of recurrent points forming diagonal lines, reflecting predictability and deterministic structure; *LAM* (Laminarity) indicates the percentage of recurrent points forming vertical lines, representing stability and resistance to change; *L* (Average Diagonal Line Length) quantifies the mean prediction time and duration of similar behavioral patterns between participants; *TT* (Trapping Time) represents the average length of vertical lines, indicating how long one participant's behavior remains in a particular state while the other changes; and *ENTR* (Entropy) reflects the complexity and diversity of diagonal line lengths, with higher values suggesting more complex interaction dynamics. Values are presented as means with 95% confidence intervals in brackets.

	Tasks	Autism Group	Typical Development
RR	Matching With Parent	45.04 [29.93, 60.15]	44.16 [26.45, 61.87]
	Convo With Parent	32.43 [10.70, 54.15]	29.24 [1.85, 56.63]
	Convo With Experimenter	29.37 [13.30, 45.44]	31.85 [13.67, 50.03]
	Aggregate Estimate	35.61 [17.73, 53.49]	35.08 [13.52, 56.64]
DET	Matching With Parent	68.64 [51.01, 86.26]	68.20 [46.10, 90.30]
	Convo With Parent	53.49 [26.24, 80.75]	46.77 [7.96, 85.59]
	Convo With Experimenter	49.31 [27.03, 71.59]	51.67 [26.09, 77.26]
	Aggregate Estimate	57.15 [34.42, 79.87]	55.55 [25.83, 85.27]
LAM	Matching With Parent	77.71 [61.62, 93.79]	78.99 [65.39, 92.60]
	Convo With Parent	66.80 [41.30, 92.29]	59.24 [23.27, 95.20]
	Convo With Experimenter	59.59 [38.03, 81.16]	60.26 [34.28, 86.24]
	Aggregate Estimate	68.03 [46.63, 89.43]	66.16 [39.37, 92.96]
L	Matching With Parent	2.82 [2.27, 3.36]	2.83 [2.21, 3.46]
	Convo With Parent	2.50 [2.00, 3.00]	2.42 [1.68, 3.17]
	Convo With Experimenter	2.44 [2.02, 2.86]	2.48 [2.03, 2.93]
	Aggregate Estimate	2.58 [2.09, 3.08]	2.58 [1.96, 3.20]
TT	Matching With Parent	3.53 [2.41, 4.65]	3.53 [2.23, 4.83]
	Convo With Parent	3.07 [2.20, 3.95]	3.15 [0.81, 5.48]
	Convo With Experimenter	2.79 [1.94, 3.64]	2.85 [2.16, 3.54]
	Aggregate Estimate	3.13 [2.18, 4.09]	3.17 [1.58, 4.77]
ENTR	Matching With Parent	1.21 [0.72, 1.71]	1.23 [0.73, 1.72]
	Convo With Parent	0.89 [0.30, 1.47]	0.75 [-0.13, 1.62]
	Convo With Experimenter	0.85 [0.38, 1.32]	0.89 [0.39, 1.39]
	Aggregate Estimate	0.98 [0.46, 1.50]	0.95 [0.31, 1.60]

S.1.7. Research Assistants

S.1.7.1 Assignment of Research Assistants

Research assistant assignments across the four testing sessions (Sessions 1, 3, 5, and 7) are illustrated in **Figure S22**. We aimed to maintain consistent RA assignments for each participant across all sessions, but this was not always feasible due to scheduling constraints and RA availability. In the autism group ($n=28$), 20 participants (71.4%) had the same RA for all four sessions, while 8 participants (28.6%) had at least one session conducted by a different RA. Similarly, in the typical development group ($n=20$), 16 participants (80.0%) maintained the same RA throughout all sessions, while 4 participants (20.0%) experienced a change in RA assignment during the study. These RA changes were documented and considered in the analysis to ensure data integrity.

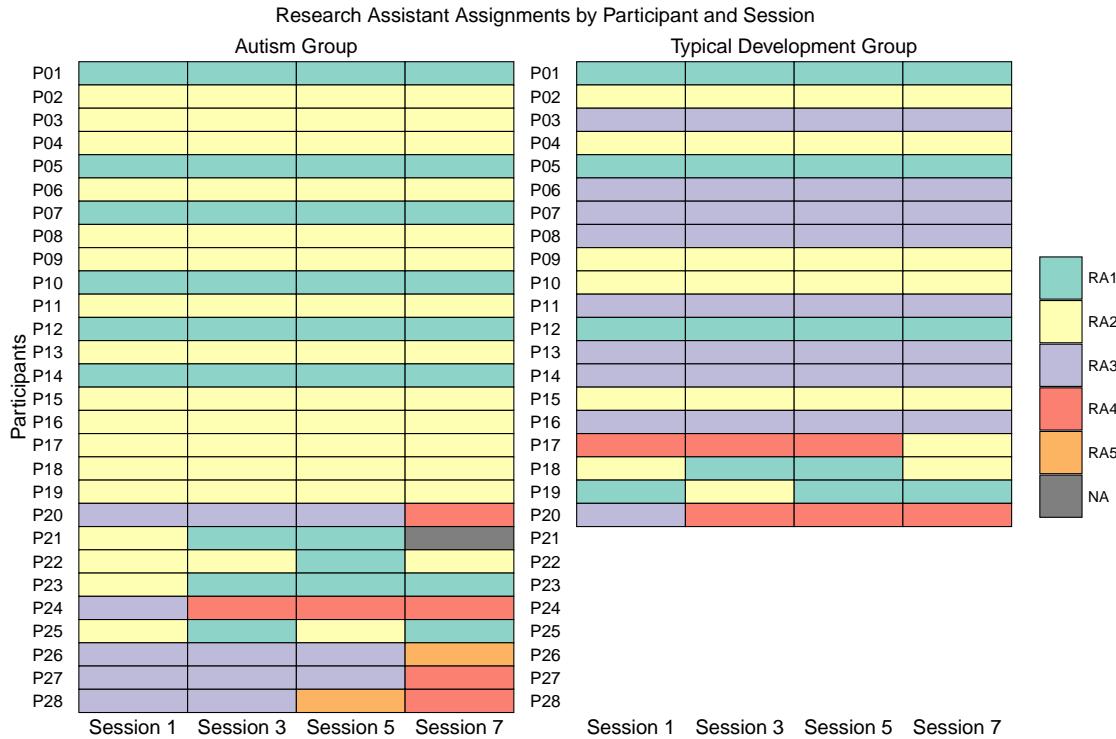


Figure S22: Each row represents an individual participant, and each column represents a testing session (Sessions 1, 3, 5, and 7). Colors indicate different research assistants (RA1-RA5). Participants are arranged by RA consistency, with those having the same RA throughout all sessions appearing first, followed by those with RA changes during the study. The missing value (NA) for participant P21 in Session 7 indicates that this session was mistakenly conducted with the parent rather than a research assistant; data from this session therefore contributed to Conversation with Parent instead.

S.1.7.2 Effects of Research Assistant

To address potential concerns about experimenter effects, we conducted a control analysis examining whether diagnostic differences in response latency were consistent across different research assistants. The diagnostic difference was consistent in both direction and rank ordering across all RAs: typically developing children showing longer response latencies than autistic children (0.23 to 0.45 seconds) (see **Figure S23**). Importantly, this effect was consistent across RAs: RAs who elicited slower responses from autistic children also elicited reliably slower responses from typically developing children, and there was no evidence of interactions between RAs and effect. This indicates that our primary findings reflect genuine

group differences rather than experimenter-specific artifacts.

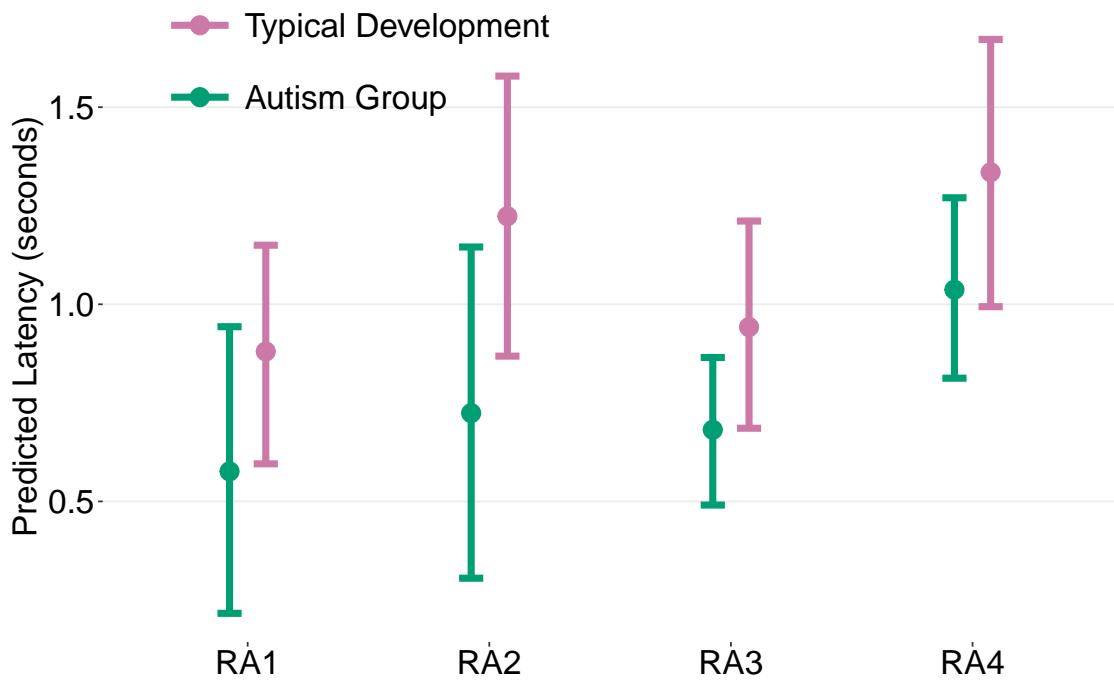


Figure S23: Response latencies by diagnostic group across research assistants. Predicted response latencies (with 95% credible intervals represented by error bars) for conversations with unfamiliar experimenters, showing consistent diagnostic differences and rank ordering across four research assistants who conducted sufficient sessions for analysis, the fifth of which only conducted two sessions and only with children in the autism group.

S.1.8. Recording Quality Assessment

All sessions were conducted via telephone using a dual-channel recording system in odd weeks and a single-channel recording system in even weeks. Technical issues were systematically documented for each session, including call connectivity problems, audio quality issues, and other disruptions that could potentially affect data collection or analysis (see **Table S15**).

Technical issues occurred in 28 of 336 total sessions (8.3%), distributed across both diagnostic groups and all testing sessions. The autism group experienced technical issues in 10 sessions (6.0% of 168 total sessions), while the typical development group encountered technical problems in 18 sessions (15.0% of 120 total sessions). The most common technical issue was temporary call disconnection followed by immediate reconnection, which occurred in 19 sessions (67.8% of all technical issues) and did not result in data loss or affect recording quality.

Recording quality was partially affected in 9 sessions (2.7% of all sessions), specifically due to audio breaking up or isolated muting events. These issues were distributed across both groups: 3 sessions in the autism group and 6 sessions in the typical development group. In all cases where recording quality was affected, the impact was deemed partial rather than complete, meaning that usable data could still be extracted from these sessions. The relatively minor technical issues had little effect on recording quality, and the timecodes showed good reliability (see **Section S.1.5**), suggesting that data collection was generally of high quality and reliable throughout the study.

Table S15: *Technical Issues by Session and Diagnostic Group.* Technical issues were systematically documented across all testing sessions for both diagnostic groups. The "Issues" column shows the number of sessions with technical problems out of the total number of sessions conducted (e.g., 4/28 indicates 4 sessions with issues out of 28 total sessions). "Affected" indicates whether the technical issue impacted recording quality: "No" indicates the issue did not affect data quality (typically brief call disconnections with immediate reconnection), while "Partial" indicates some impact on audio quality that could potentially affect analysis (e.g., audio breakup, persistent muting events). Details specify the type and frequency of issues: Call dropped = temporary disconnection with reconnection; Breakup = partial audio quality degradation; Muting = isolated audio muting events. ASD = Autism Spectrum Disorder group; TD = Typical Development group.

	Session	Issues	Affected	Details
ASD	1	4/28	Partial	Call dropped (2); Breakup (1); Muting (1)
	2	0/28	No	No issues
	3	1/28	No	Call dropped (1)
	4	0/28	No	No issues
	5	2/28	No	Call dropped (2)
	6	0/28	No	No issues
	7	3/28	Partial	Call dropped (2); Breakup (1)
TD	1	5/20	No	Call dropped (5)
	2	3/20	No	Call dropped (3)
	3	3/20	Partial	Call dropped (1); Breakup (1); Muting (1)
	4	1/20	No	Call dropped (1)
	5	2/20	Partial	Call dropped (1); Muting (1)
	6	1/20	No	Call dropped (1)
	7	3/20	Partial	Muting (3)