technische universität dortmund

reha wiss Fakultät Rehabilitationswissenschaften

Predicting later advanced language and metacognitive skills from initiations of joint interaction and communicative goals in earlier development

vorgelegt von

Camilla Crawshaw

als Dissertation zur Erlangung des Grades einer Doktorin

der Philosophie (Dr. phil.)

in der

Fakultät Rehabilitationswissenschaften

der Technischen Universität Dortmund

Dortmund

2025

Betreuerin: Prof. Dr. Ute Ritterfeld

Betreuerin: Prof. Dr. Carina Lüke

Die vorliegende Arbeit wurde von der Fakultät Rehabilitationswissenschaften der

Technischen Universität Dortmund als Dissertation angenommen.

Gutachterin: Prof. Dr. Ute Ritterfeld

Gutachterin: Prof. Dr. Carina Lüke

Tag der Disputation: 06.10.2025

Dortmund

Acknowledgements

Firstly and foremostly, I wish to thank my first supervisor Ute Ritterfeld and my second supervisor Carina Lüke for their constant support, patience, and encouragement over these many years. I have learned and grown so much through their mentoring while their continued belief in me has inspired and enabled me to persist with this work, for which I shall be forever grateful.

Secondly, I would like to thank every member of the S&K department at TU Dortmund (all my colleagues, our students and our research assistants) for their feedback, insights, discussions, and practical support. I particularly thank Jannika Böse and Annika Büssing for their many hours involved in carrying out the data collection as well as the preparation and administration associated with it. I am also very thankful for the many hours of support with coding and transcription provided by Olivia Espeter and Tabea Musshoff. I am also very grateful to everyone else who has ever contributed to, collaborated on or worked within previous stages of the wider longitudinal project in which I carried out this doctoral work.

Thirdly, I wish to extend a huge thank you to every family and each child who has participated in this longitudinal study again and again over so many years. Without their time, commitment, and willingness, this work would never have been possible.

Finally, I thank my own extended family, my daughter Ottilia for giving me the motivation to persevere, and particularly Gareth, without whose all-encompassing support I never could have finished this work.

Table of Contents

Abstract

Gesture is an important early precursor to spoken language development but also continues to co-develop alongside it within an integrated communication system. As a multimodal, embodied form of communication, gestures allow us to refer to, represent, and express diverse information in a visuospatial format for our interlocutors. Prior work has shown that early use of gesture (e.g. to initiate joint interaction and establish communicative goals) can predict later language skills, which in turn have separately been found to predict Theory of Mind (ToM) ability in addition to having been linked to performance on metaphor comprehension tasks. Children with language delay have also been shown to exhibit delays in gestural development. However, it had not yet been well identified (1) whether early use and processing of gesture could also predict later language-associated, metacognitive and metalinguistic skills such as ToM or metaphor comprehension, or (2) what such a connection might look like in atypically developing children with a history of language delay.

Within the context of a wider longitudinal study spanning the ages of 12 months and 9 years, 35 children's gestural, spoken language, ToM, and metaphor comprehension abilities were investigated and analysed. At age 9, children's performances across these areas were also analysed cross-sectionally on the basis of whether or not they had a history of language delay. Results revealed that a child's ability to point specifically with the index-finger at age 12 months could directly predict their performance at age 9 years on a metaphor comprehension task. Both iconic gesture comprehension and language skills at age 3 years were also found to mediate an indirect relation between index-finger pointing at age 12 months and ToM and metaphor comprehension performance at age 9 years. Having a Late Talker status at age 2 years also appeared to influence performance across certain but not all language measures, one advanced ToM task and a metaphor comprehension task. These findings indicate that early gesture production and comprehension skills may support and facilitate not only children's developing language skills but also their capacities for ToM and metaphor comprehension.

Publications Supporting the Doctoral Dissertation

This cumulative thesis *Predicting later advanced language and metacognitive skills from initiations of joint interaction and communicative goals in earlier development* covers the works by Crawshaw et al. (2025a, 2025b) but not those by Crawshaw et al. (2020), Tolksdorf, Crawshaw et al. (2021) and Tolksdorf, Viertel et al. (2021) also carried out during this author's doctoral studies. The two works supporting this dissertation have been published in international, peer-reviewed journals:

Publication I (Appendix A):	Crawshaw, C. E., Lüke, C., & Ritterfeld, U. (2025b). Does early gesture usage contribute alongside oral language to later theory of mind performance and metaphor comprehension? Indications from a longitudinal study with children aged 1–9 years. *Language Acquisition*, 1–31. https://doi.org/10.1080/10489223.2025.2455174
Publication II (Appendix B):	Crawshaw, C. E., Lüke, C., & Ritterfeld, U. (2025a). Late Talkers' Language, Metaphor, Theory of Mind, and Reading Skills at 9 Years. *Journal of Speech, Language, and Hearing Research*, *68*(3), 1038–1055. https://doi.org/10.1044/2024_JSLHR-23-00703

List of Tables and Figures

Tables

Figures

X

1 Introduction

Language and communication are shaped, formed, and facilitated by many more aspects and factors than the sounds we can produce. Co-speech gestures—bodily movements we make to support verbal exchanges—are a fundamental building block of interpersonal interaction and may have played a key role in the evolution of spoken language (Tomasello, 2019). Although we can make some gestures with other bodily parts and facial expressions are also extremely important for communication, our hands afford us a more sophisticated modality in referring to and representing external entities, concepts, and events. Producing and comprehending gestures inherently draws on and requires integration of visual and mental perception as well as physical and cognitive coordination. This requires motor, representational, inferential, and decoding skills along with a recursive understanding of intentionality and conventionalisation. All of these competencies are also relevant for general and advanced language as well as metacognitive development. Index-finger pointing presents an early precursor and form of conventionalised gesture, whereby infants must coordinate their gaze with that of an interlocutor to access a joint referent for their attention. Seeing someone point to something, draws your attention to that referent but it also instigates a process of considering why they pointed to it: What was the intended purpose of that gesture? Similarly, the pointer themselves must create a gesture that is easily comprehendible and unambiguously singles out a referent (e.g. pointing clearly with the index-finger instead of vaguely waving a hand in a given direction). This requires thinking about how the viewer of the point will perceive it. As the pointer, your goal is to direct the viewer’s attention to a referent with the possible intention that the viewer will begin to ask themselves why you have referred to it, creating a common-ground, joint communicative context. Thinking about others’ goals, intentions, objectives—the contents of their minds—is a starting point for Theory of Mind (ToM).

ToM refers to the ability to impute mental states to oneself and others (Premack & Woodruff, 1978). Mental states include purpose and intention but also many others such as knowledge, (false) belief, thought, emotion, desires, doubt, estimation, pretence, likes and dislikes, and so forth. ToM draws on visual, affective, and cognitive perspective-taking skills as well as the abilities to follow and coordinate gaze, attention, and goal orientation. It also necessitates an understanding of (mis- and meta-) representation and making distinctions between appearance and reality. ToM is crucially important for effective and successful social interaction. Human development takes place across a protracted time span and occurs within a highly cooperative social environment, whereby infants face unique socio-cognitive challenges in order to secure care and attention from a range of adults (Tomasello, 2019, p.25). As children grow older, their social world expands to peers and they develop an understanding of collective intentionality, socio-cultural norms and conventions as well as a sense of fairness (Tomasello, 2019). Children also begin to monitor their actions and thinking based on their social group members' perspectives and evaluations (Tomasello, 2019). Without a functioning ToM, it would be extremely difficult for children to successfully access and participate within society.

Sometimes our use of language may not be immediately transparent to our interlocutors. Nonliteral, figurative language such as metaphor requires us to use further semantic, pragmatic, inferential, and metalinguistic skills to decode a speaker's intended meaning. When using metaphor, we make conceptual connections between separate entities or events by applying a word or phrase in a nonliteral but representative or symbolic way. In order to access this intended meaning we need an understanding of representations and may require perspective-taking skills (i.e. ToM). While some metaphors can become conventionalised over time and their processing may become automatised, humans communicate creatively and novel metaphors still emerge. Being able to navigate complex, nonliteral language becomes increasingly important for children to successfully interact with their peers and for their academic achievement within a school setting. Like metaphors, gestures can also represent (or analogically refer to) abstract concepts, concrete entities or events. Embodying language and communication, gestures can also be metaphoric in themselves (McNeill, 1992) permitting us to contextually embed a visuospatial depiction of a metaphor as well as a linguistic one. If we are able to process them appropriately, representational gestures and metaphors both augment and support communicative exchanges, furnishing us with a window into others' minds.

However, when it comes to language and socio-cognitive skills, we are not all equally endowed and do not follow identical developmental trajectories. What is easy and self-explanatory for one person may not be for another. We are all individuals and children vary widely in where they fall along various developmental spectra (Bates et al., 1995). Children who develop atypically and follow divergent developmental trajectories may need extra support with certain skills to thrive within society, successfully navigate peer interaction, effectively participate in discourse, as well as achieve and progress at school. Intervention may be necessary or helpful for late talkers (LTs) who may or may not go on to receive a diagnosis of developmental language disorder (DLD), children on the autism spectrum, those who experience other neurodevelopmental disorders (e.g. Down's Syndrome, William's Syndrome, schizophrenia), or physical challenges such as children who are visually impaired, deaf or hard-of-hearing. Research has established that metaphor and ToM acquisition can be delayed, disrupted, or disordered to varying degrees across atypically developing populations where language has been implicated (e.g. Bühler et al., 2018; Kalandadze et al., 2018; Nilsson & de Lopez, 2016; Smit et al., 2019; Vissers & Koolen, 2016; for a discussion, see Appendix B: Crawshaw et al., 2025a). Gestural development also transpires differently within populations experiencing language delay, varying in terms of delay or deficit depending on both the clinical population and the gesture type (e.g. Capone & McGregor, 2004). Surprisingly, few others have considered the implications of early gesture-speech integration for the later development of language-related skills such as ToM and metaphor comprehension (c.f. Appendix A: Crawshaw et al., 2025b).

The following chapters will address each of these interrelated areas of development in more detail: *Gesture and Spoken Language as an Integrated System* (Chapter 2), *Theory* *of Mind* (Chapter 3), *Metaphor Comprehension* (Chapter 4). These competencies will be brought together and discussed across the subsequent chapters. The overarching research question and an overview of the two publications supporting this cumulative dissertation will first be presented in Chapter 5: *Overarching Research Question and Overview of Supporting Publications*. An overall discussion of these publications and a reflection upon their limitations will then be provided in Chapter 6: *Overall Discussion*. Final conclusions and future directions will be presented in Chapter 7: *Conclusions and Future Directions*.

2 Gesture and Spoken Language as an Integrated System

If research is focused solely on a single mode of communication, i.e. spoken language, then the ensuing findings do not reflect the full, multimodal picture of language acquisition. More holistic approaches can give us greater insights into these developmental processes. Research has demonstrated that gesture and spoken language form a closely integrated communication system (e.g. Capone & McGregor, 2004; Goldin-Meadow, 1998; Graziano & Gullberg, 2018; Holler & Levinson, 2019; Kelly et al., 2010; Kendon, 2004; Kita & Özyürek, 2003; Lüke et al., 2020; McNeill, 1992, 1985; Özyürek, 2014; for a comprehensive review, see Wagner et al., 2014). Co-speech gesture is crucially important for effective and successful communication in diverse ways. Gestures can help reinforce co-occurring speech visually (complementary) or provide additional (supplementary) information which has not been conveyed verbally, as well as offering a substitute for and an alternative route of communication (e.g. Kendon, 2004). Using gestures improves communicative efficiency, allowing us to convey visual information about spatial dimensions and bypass the need for precise verbal descriptions (Wagner et al., 2014).

Gestures can also serve prosodic and pragmatic functions, modulating the rhythm of speech, signalling levels of attention and conversational turn-taking, as well as organising discourse (Kendon, 2004; McNeill, 1992; Wagner et al., 2014). These types of gestures include ‘beat gestures’, which convey no propositional content of their own but perform discourse-oriented functions such as indicating the importance of linguistic inter-item relations or highlighting extra-narrative discontinuity in the discourse structure (McNeill, 1985, p.359). Repairing instances of miscommunication, gestures can also occur while speakers struggle to conduct lexical recall or structure sentences appropriately (McNeill, 1992). Research has also suggested that use of gesture serves a cognitive function for the speaker, priming lexical retrieval by maintaining visual and spatial imagery (Wesp et al., 2001) and supporting narrative recall when produced during encoding (Sweller et al., 2024). People still use gestures even when not these are not visible to their interlocutor and speakers will actually gesture less frequently when an object being described is visible to both interlocutors (Wesp et al., 2001). Co-speech gesture is thus a highly multifaceted communicative and cognitive tool.

2.1 Gesture: Early Beginnings of Language

Gesture partially predates verbal language: Deictic (pointing) gestures emerge preverbally while later types of gesture such as iconic (representational) gestures emerge alongside oral language (c.f. Bates et al., 1975, 1989; Capone & McGregor, 2004; Guidetti & Nicoladis, 2008). Iconic gestures are a type of representational gesture which depict aspects of a verbally described event or situation through their form or manner of execution (McNeill & Levy, 1982, p.275). These gestures can also portray and embed different perspectives or viewpoints (McNeill & Levy, 1982; Parrill et al., 2018): (1) Character viewpoint gestures embody agents or characters to depict their actions from a first-person perspective; (2) observer viewpoint gestures depict objects or events from a third-person perspective; (3) sometimes a gesture can combine character and observer viewpoints to take a dual perspective; and (4) some iconic gestures possess no viewpoint and merely trace a shape. Character viewpoint gestures often occur in children's narratives and retellings and are associated with higher narrative structure scores (Parrill et al., 2018). Narrative comprehension and production skills have importantly also been linked to ToM abilities (e.g. Lorusso et al., 2007; Kim et al., 2021); these abilities are arguably utilised when performing gestures from a character's perspective during a narrative (Demir et al., 2015). To be able to creatively produce iconic gestures, children must draw on a certain level of imitation, mental simulation, imagination, symbolic representation, or pretence skills to abstract from and communicate something relevant to an absent action (Tomasello, 2019, p.108). In this way, gestures are representational actions, providing meaningful substitutions and functioning as analogical stand-ins for various ideas, objects, actions, and relations (Novack & Goldin-Meadow, 2017, p.652). Another interesting and relevant subtype of iconic gestures are metaphoric gestures, which refer to abstract concepts rather than more concrete ones: The form of the gesture indirectly conveys meaning by iconically depicting the vehicle of a metaphor (McNeill & Levy, 1982). These gestures are even further distanced from the original absent action and related concept.

Gesture appears to be a universal, language-independent, prelinguistic basis for human communication (Liszkowski et al., 2012). Liszkowski et al. (2012) conducted a cross-linguistic and cross-cultural study, using a standardised, semi-natural elicitation procedure to investigate infant preverbal pointing (deictic gestures). By 10–14 months old within the same experimental situation, both infants and their caregivers across all cultures pointed with similar frequencies as well as using the same proto-typical extended index finger (Liszkowski et al., 2012). Both deictic and iconic gestures have been observed to mediate early language acquisition trajectories across typically and atypically developing children (Botting et al., 2010; Capone & McGregor, 2004; Colonnesi et al., 2010; Goldin-Meadow, 2020; Iverson et al., 2018; Lüke et al., 2020; Mainela-Arnold et al., 2014; Rowe & Goldin-Meadow, 2009; Wray et al., 2016; however see Kirk et al., 2022 for a more conservative meta-analysis).

While infants and children are still limited in the range of ideas they can express verbally, gesturing allows them to extend their communicative scope (Goldin-Meadow, 1998). Early use of gesture predicts later spoken language ability: It is associated with advances in both expressive and receptive language (Capone & McGregor, 2004); cross-modal combinations support the transition to two-word speech (Capone & McGregor, 2004; Goldin-Meadow, 1998); and increased parental use of gesture-word pairs facilitates children's earlier acquisition of symbols (Capone & McGregor, 2004). Even for very young children, observing gestures benefits comprehension of accompanying spoken language in many different ways, e.g. directing attention, emphasising important points, integrating and reinforcing semantic information (for a meta-analysis, see Dargue et al., 2019). The findings of Dargue et al.'s (2019) meta-analysis demonstrated improved comprehension when observing deictic, iconic, and metaphoric gestures, as well as a mixture of these. Gestures may thus provide pragmatic direction in supporting word learning (Falkum, 2019). Clearly, gesture use yields advantages for both the speaker and listener across the age span, in supporting development, and within a wide variety of contexts (c.f. Capone & McGregor, 2004; Dargue et al., 2019).

Developmental relations between gesture and spoken language can however follow different pathways and trajectories within atypically developing populations such as late talkers or those experiencing DLD (e.g. Capone & McGregor, 2004; Guidetti & Nicoladis, 2008). Early on, infants and toddlers with language delay produce fewer gestures than their typically developing peers, while those with an additional diagnosis of autism produce even fewer (c.f. Iverson et al., 2018; Lüke, Ritterfeld et al., 2017; Manwaring et al., 2019; Sansavini et al., 2021; Thal et al., 2013; Watson et al., 2013). Later on this trajectory appears to change course and children with DLD begin to use gesture more frequently than their typically developing peers albeit in a compensatory manner to replace words they are not able to verbalise (Lüke, Ritterfeld et al., 2017; Mainela-Arnold et al., 2014; Wray et al., 2017). Children with DLD also perform more poorly on measures of gesture comprehension and show weaknesses in gesture accuracy (Wray et al., 2016, 2017). However, just as in the case of typically developing children, children experiencing language disorders can harness gesture to communicate when they do not have fully developed articulatory and language systems (Capone & McGregor, 2004, p.180). If their gestures are recognised as communication attempts, they can further elicit language learning opportunities from caregivers around them (Capone & McGregor, 2004). These findings emphasise the adaptive, closely intertwined nature of the relation between gesture and spoken language continuing across the development span.

2.2 Gesture as a Developmental Scaffold

Overall language outcomes have been associated with how well intentional communication is understood (Camaioni, 1992). As an example of something well understood among shared communicators, deictic pointing with specifically the index-finger can be identified as a precursor to communicative conventionalisation (Liszkowski & Tomasello, 2011). By providing a nonverbal method for establishing joint attention and forming joint intentions, it may mark a fundamental turning point in children's early socio-cognitive understanding of communicative intentions and shared intentionality (Liszkowski & Tomasello, 2011). Through observing the developmental sequence of deictic gesturing in children (showing off, showing objects, giving, pointing-for-self, and pointing-for-others), we can witness a gradually distancing of 'self' from 'object' and 'other', necessary for further symbolic development (Bates et al., 1975, pp.217–218; Capone & McGregor, 2004, p.175). Bates and colleagues cited Werner and Kaplan's (1964, as cited in Bates et al., 1975, p.217–218) argument that pointing serves a cognitive function in helping a child to highlight the distinction between 'self' and 'object'. In doing so, the authors proposed that pointing-for-self first facilitates an important cognitive step in children's development of the concept of reference by allowing them to progressively distance themselves from an object-of-reference (Bates et al., 1975). This distancing process might support children's emerging thinking about relations between their internal selves, external objects or entities, and time or space. Bates and colleagues suggested that pointing might then be later separately integrated into communicative schemes once this concept of reference is more stably developed (1975). The above works demonstrate that gestural development reflects both cognitive and communicative advances.

Early pointing production and comprehension evidently supports language acquisition (Capone & McGregor, 2004) and infants use pointing to communicate meaningfully in diverse ways (c.f. Liszkowski et al., 2012). Prior work has observed two pathways of progression in children's deictic gesture development. Firstly, children undergo a transition from whole-hand to index-finger pointing between 12–21 months (Lüke, 2015; Lüke, Ritterfeld et al., 2017). Secondly, children's production and understanding of communicative motivations for pointing advance from imperative to declarative-expressive to declarative-informative motives (Camaioni et al., 2004; Lüke, 2015; Lüke, Grimminger et al., 2017; Rohlfing et al., 2022). Production of declarative pointing has also been linked to an understanding of others' intentions (Camaioni et al., 2004; Liszkowski & Tomasello, 2011). However, the handshape of the point (i.e. specifically using the index-finger instead of gesturing more ambiguously with the whole-hand) appears to be qualitatively more important than any particular motive for predicting spoken language ability (Lüke, Grimminger et al., 2017). This may be because all motives for pointing (expressive, imperative, informative) are cooperative and engage in shared intentionality to direct attention to or share emotions and attitudes about external entities and situations (Tomasello, 2019, p.98). Index-finger points enable a speaker to single something out for their interlocutor much more clearly than by waving a hand vaguely in a general direction. Their conventionalised status also helps underscore a speaker's intention to refer to something. Through a speaker's understanding and exploitation of this conventionalisation to design a gesture for their interlocutor, it may then straddle deictic and representational bounds. Three features of gesture may therefore help scaffold children's developing socio-cognitive understanding of communicative representations: (1) index-finger pointing signifies something distinctively in communication, (2) interlocutors may differentially perceive variants of gestures, and thus (3) gestures can be designed to better support an interlocutor's comprehension within a communicative context. In integrating such entailments, children are tapping into developing ToM skills.

Analysing how and when developmental processes transpire atypically often facilitates greater understanding of how they take place within typical development. Longitudinal work by Thal and colleagues has highlighted how early gesture use reflects the ability to represent knowledge symbolically, offering insight into LTs' representational abilities when linked with early language skills (Thal et al., 2013, p.173). Their work identified three groups of children at 16 months of age: (1) those who were typically developing, (2) those with delayed expressive language but receptive language in the normal range ('late producers'), and (3) those with both delayed expressive and receptive language ('late comprehenders'); group differences in both receptive vocabulary and use of gesture could already be observed at 10 months of age (Thal et al., 2013, p.179–180). Typically developing children used significantly more gestures (both early communicative, non-symbolic ones as well as later representational types) than ‘late producers’, with ‘late comprehenders’ using the fewest (Thal et al., 2013, p. 179–180). These group trends were also reflected across a wide variety of language measures, persisting even at age 7 years (Thal et al., 2013, p.186). Use of gestural symbols preceded spoken verbal symbols and predicted language outcomes a year later: ‘late producers’ with sufficient use of representational gestures and receptive language in the normal range were more likely to ‘recover’, while ‘late comprehenders’ who infrequently used communicative gestures remained significantly delayed (Thal et al., 2013, p.192), i.e. those who might go on to receive a diagnosis of DLD. Thal et al. concluded that ‘late producers’ use of gesture reflected a higher level of representational ability than that of ‘late comprehenders’ (2013, p.192). This work underlines the important potential for gesture to scaffold advances in representational ability in both typically and atypically developing populations.

Gesture use may also support conceptual development, functioning as a scaffold in situations requiring children to comprehend decontextualised symbols (Capone & McGregor, 2004, p.178). Children gradually become more sophisticated in their use of symbolic gesture forms. The transition from gestures which use a body part to concretely depict an object to gestures which require both the gesturer and the viewer to abstractly imagine an object, reflects a further gradual distancing of children’s symbolic representations between substitute and referent object (Boyatsis & Watson, 1993, p.734). The supportive nature of gesture in children’s abstract concept acquisition has also been demonstrated at later ages in the context of mathematics at school (c.f. Goldin-Meadow, 2000: mathematical concepts such as quotients, factors, and even limits in calculus). Gesture-speech mismatch also reveals information about children’s knowledge and underlying mental representations while these are not yet stable, e.g. a child may incorrectly verbally describe a given mathematical concept while their gesture correctly depicts it (Goldin-Meadow, 2000). Educators can use gestures to improve children's acquisition of concepts through visual reinforcement but can also read children's gestures to better assess their level of understanding and consequently adapt teaching methods (Goldin-Meadow, 2000; Capone & McGregor, 2004). This demonstrates the potential power of gesture for prospective intervention contexts.

The work reviewed in this section critically emphasises the importance of gesture in representational and conceptual development, underpinning and facilitating language development. This early relation is vital to carry forward when considering other language-related areas which draw on metarepresentational, metacognitive, and metalinguistic skills, such as ToM and metaphor comprehension (c.f. Crawshaw et al., 2025a, 2025b; Appendices A and B).

2.3 Gesture, Embodiment and Embeddedness

Cognitive and communicative skills in humans have evolved by necessity from interactions with and manipulations of the physical world: e.g. spatial awareness to locate food, categorising objects to identify food, distinguishing quantities to maximise food intake (Tomasello, 2019, p.12). Causal inferencing or thinking may have developed from the need to cognitively represent a problem and mentally simulate potential uses of available tools in order to select a tool in a different location from the context of the problem (Tomasello, 2019, p.12). In this respect, the thinker is cognitively embedding an internal, embodied representation of themselves within an external situation or context to solve a problem. Embodied cognition theories posit that our bodily properties inform and constrain how we perceive and conceptualise our surroundings, underpinning our ability to form mental representations (c.f. Hostetter & Alibali, 2008). Embedded cognition theories propose that we then navigate such external contexts and environments by cognitively embedding and mentally simulating these internal, embodied representations of ourselves within them (e.g. Huebner, 2013). It has been proposed that gestures may stem from embodied simulations of motor and perceptual states occurring during speech and thought (Hostetter & Alibali, 2019, 2008; for a discussion, see Appendix A: Crawshaw et al., 2025b). As a visible depiction or 'embodiment' of language which can directly express spatial and motor information, their use facilitates speech about mental images, allowing communicators to efficiently circumvent more difficult verbal encodings (Hostetter & Alibali, 2008).

According to embodied cognition theories, connecting words to real-world referents and simulating their perceptual and motor properties allows us to process language about both physically present and concrete absent referents as well as abstract concepts (c.f. Hostetter & Alibali, 2008, p.498; Pecher et al., 2011). Similarly, McNeill and Levy (1982, p.271–272) proposed that language is generated from a conceptual basis, with abstract ideas represented linguistically via concrete models of reality and gestures visually revealing and extending these concrete models for direct observation. Sensory-motor images or image schemas form these physical and spatial concrete models of reality; we then map abstract concepts onto our understanding of them (c.f. McNeill & Levy, 1982, p.292; Hostetter & Alibali, 2008, p.498; Pecher et al., 2011). It has also been suggested that this mapping process is metaphorical in nature, given that gesture forms have been found to be visually consistent with common families of linguistic metaphors (McNeill & Levy, 1982, p.289–291; see also Cienki & Müller, 2008). Lakoff and Johnson (1980) argue that language is based on a metaphorical conceptual system, supported by a large network connecting various literal, metaphorical, conceptual, perspectual, and aspectual entailments. For example, when we talk about 'theories', we conform with a family of metaphors which conceptualise them as 'buildings': They are weak or strong, constructed, can be supported or bolstered with solid arguments, have a basis or foundation, possess a framework, can stand up or fall down and collapse when faced with criticism (Lakoff & Johnson, 1980). In this way, our reality is structured by metaphorical entailments which help us to effectively and coherently conceptualise and communicate aspects of our experience within a conventionalised, 'common-ground' system (Lakoff & Johnson, 1980). This common ground could underpin our thinking about others' intentionality, perspectives, and contexts of joint interaction.

As an embodiment of language, gesture coordinates with speech to concretely depict an object or event described simultaneously in a verbal format, drawing on a specific cognitive representation to make a single, semantically coherent reference (McNeill, 1985; Goldin-Meadow, 1998). Going beyond the idea of simulated action, Novack and Goldin-Meadow (2017, p.653) argued that gestures should also be viewed as representational actions, given that all types of gesture abstract away from their physical bodily movement to represent some other communicative intent. Combining ideas about embodied and embedded cognition, it is possible that gestures first embody a simulated action and are then embedded within communication to convey it representationally.

Interestingly, prior work has demonstrated that the types of gestures used by speakers vary depending upon the speech they accompany, drawing on multiple representation systems (e.g. Zdrazilova et al., 2018). Zdrazilova and colleagues (2018) found that participants used more metaphorical and beat gestures when speaking about abstract word meanings and more iconic gestures when talking about concrete word meanings. This was parallelled in spoken language; abstract word meanings were communicated with more references to people and introspections while concrete word meanings were communicated with greater reference to objects and entities (Zdrazilova et al., 2018). However, gesture types are not always restrictive and can sometimes straddle categories. Speakers can use metaphoric gestures to better convey complex representations and relations through the presentation of various kinds of supporting information, including mental states, mathematical concepts, metaphysical phenomena, spatial imagery, and so forth (Wagner et al., 2014, p.215). Beat gestures can convey complementary abstract meaning regarding dialogue structure, referring to synchronised words or different points within a conversation (Wagner et al., 2014, p.215; see also McNeill, 1992). Pointing gestures also co-occur with topic shifts, referring to contrasting spaces to signal a transition from a previous topic to the next (Wagner et al., 2014, p.215; see also McNeill, 1992). The main interactive functions of the above gestures are thus mixed with complementary, abstract or metaphorical references (Wagner et al., 2014, p.215). In this way, they may be analogically representing relations and expressing similarities between certain common-ground conceptual models of reality (e.g. those discussed by Lakoff & Johnson, 1980; McNeill & Levy, 1982).

Embodying and visually depicting relations between mental image schema and linguistic representations, gestures convey representational content through being embedded within a communicative context. However, communicative contexts are built with socio-cognitive foundations, relying upon concepts such as shared intentionality. These concepts will be discussed further, with a return to the role of gesture, in the following chapter: *Theory of Mind*.

3 Theory of Mind

Being able to represent and reflect upon the contents of our own and others' minds (perceptions, goals, attitudes, emotions, evaluations, knowledge, beliefs, and so forth) is vitally important for our existence within a fundamentally cooperative society. Theory of Mind (ToM) enables us to recognise, integrate, exert influence over, and act in response to (or anticipation of) the mental states of ourselves and those around us. Children learn that our minds actively construct representations of reality instead of only mirroring it (Miller, 2006).

Human ToM has evolved over time, originating from the concept of intentional agency in competitive foraging contexts, e.g. identifying the goals and perceptions of competitors to predict their behaviour (Tomasello, 2019, p.12). This would have necessitated an understanding of visual perception, with competitors holding an advantage if their thoughts remained unknown (Tomasello, 2019). However, human ToM extends mind-reading from competitive to cooperative purposes, requiring recursive social and mental coordination of self and other: I want you to also know my thoughts (Tomasello, 2019, p.43). In this way, shared intentionality goes beyond individual intentionality, we share attention and goals, communicate, and participate in others' thinking, rather than simply comprehending, predicting and manipulating our physical or social worlds (Tomasello, 2019, p.13). We also self-monitor and regulate our actions and cognition on the basis of our social group members' perspectives and evaluations (Tomasello, 2019, p.14).

Through attempting to perceive and understand the world with another person while maintaining our own perspective, we create "perspectival cognitive representations": abstracting representations to simultaneously perceive an entity from different perspectives under different descriptions (Tomasello, 2019, p.16). Engaging in joint intentional activities and sharing conceptual worlds likely provided the pragmatic foundations for communication and language skills (Tomasello, 2019, p.16). Within these shared conceptual worlds, gestures may have marked a form of continuity during the transition from sensed meaning to emerging 'conscious' meaning, originating as natural signs but later supporting association between a stimulus and mental content (Baggio, 2025, p.19). Before having access to spoken language, humans would have begun to use pointing and ‘pantomiming’ (representational) gestures to coordinate their perspectives and intended roles in pursuing a collaborative goal (Tomasello, 2019, p.16). Thus, gesture would have served as an important prelinguistic medium for conveying, communicating, and influencing early ToM. Research has correspondingly demonstrated that gesture production is sensitive to ToM: Adults modify their gestures to meet the needs of a recipient in emphasising relevant features of targets (Kim & Schachner, 2020).

To make a socially recursive inference about a communicator’s gesture, the viewer is required to conceptually embed an intentional or mental state within another (Tomasello, 2019, p.16). In this way, use of pointing gestures would have scaffolded and guided developing cognitive representational skills. Infants begin by engaging in acts of joint attention, through which they become capable of conceptualising entities and situations from different perspectives simultaneously, before finally being able to view things from an ‘objective’ perspective (Tomasello, 2019, p.44). They learn to coordinate gaze and ostensively point, while continuing to monitor the attention and response of the adult with whom they are interacting. These referential communicative contexts require recursive inference-making about embedded mental states, facilitating their entrance into a world of shared linguistic conventions (Tomasello, 2019, p.44). Early communication about mental states thus begins with gesture before spoken language takes the reins (See Appendix A: Crawshaw et al., 2025b for further consideration of the links between both deictic and iconic gesture and ToM).

Language shapes our thinking and interactions. It functions as both an intra-individual representational system and as an inter-individual communicative system (Astington & Baird, 2005). Shared linguistic conventions coordinate social activities and conventionalise perspectives by allowing us to concurrently attend to, reflect upon, and exchange ideas (Tomasello, 2019, p.20). For example, pluralisation enables us to express generic propositions about the world and build ‘objective’ perspectives (Tomasello, 2019, p.20): e.g. “Books can be read.” Mental or internal state language provides a semantic and syntactic framework for us to represent desire, belief, knowledge, hearsay, perception, emotion, and so on (c.f. Astington and Baird, 2005). Children learn very early on that the intentional verb ‘want’ takes an irrealis complement of either a thing or a proposition (e.g. “I want the toy,” or “I want to be held by you,”) requiring them to make a mapping between the word, behavioural manifestations of desire and a projected intentional state (de Villiers, 2007). Complement clause structures allow us to represent and embed such (recursive) propositions about other’s (false) beliefs or mental states, distancing ourselves from the verity of their content (de Villiers, 2007; Durrleman, 2020): e.g. “Mary thought that trees grew on clouds,” or “John said that Mary felt that colours could be heard.” In this way, these structures may furnish us with the representational adequacy to both reason about and convey the potentially false content of other’s minds (de Villiers, 2007). Interestingly, existing work has shown that language-impaired or -delayed populations struggling with grammatical skills such as complementation perform more poorly on ToM tasks (c.f. Andrés-Roqueta et al., 2013; Durrleman, 2020; de Villiers, 2007; for a more detailed discussion of ToM development in LTs and DLD, see Appendix B: Crawshaw et al., 2025a). Through spoken language, we formulate, express, and process complex thoughts, emotions, perspectives; it is thereby our medium for communicating and understanding ToM.

A large body of work has consistently demonstrated bidirectional predictive relations between language and ToM development (e.g. de Villiers, 2007; Ebert, 2020; meta-analysis by Milligan et al., 2007; review by Miller, 2006; for further discussion including a description of commonly used tasks, see Appendix A: Crawshaw et al., 2025b). Some have also argued that ToM development is causally dependent on a convergence of adequate language and social experience, whereby children’s conversations and interactions scaffold and facilitate their growing social understanding (Garfield et al., 2001; see also de Rosnay & Hughes, 2006). However a certain level of intentionality (early ToM) and inferencing skills may likewise be necessary for acquiring language (Miller, 2006; Papafragou, 2001; Tomasello, 2019). There are instances where a certain level of ToM might first need to be reached before particular aspects of language may later be acquired and mapped. These include modality (permission, obligation, ability, intention, contextual inference: e.g. ‘may’, ‘should’, ‘must’, ‘have to’, ‘can’, ‘will’ [+be]) and evidentiality (the speaker’s assessment of their informational source in terms of memory, observation, hearsay, inference, etc. as well as their degree of certainty: strong vs. weak) across a range of languages (Papafragou, 2001). Still, language and ToM development are so tightly interwoven that it is difficult to tease them apart; even infants’ early participation in joint attention and their acquisition of intentionality occur within the context of communicative acts (Miller, 2006). ToM development itself is also a non-linear process spanning many different aspects and constructs. While young children transition with conceptual continuity from implicit to explicit to advanced ToM (Sodian et al., 2020; Osterhaus et al., 2022), longitudinal research has also demonstrated weak relations between early and advanced ToM performance, with substantial progress made well beyond preschool age (Osterhaus & Koerber, 2021a, 2021b). ToM has also been linked to other abilities such as executive function including self-regulation which are additionally relevant for language development (c.f. Weimer et al., 2021).

Prior work attempting to specify exactly which areas of language may (uni- or bidirectionally) predict ToM abilities have also produced conflicting results both within and across languages (for a wider discussion, see Crawshaw et al., 2025a, 2025b; Appendices A and B). For example, De Mulder et al. (2019) worked with Dutch-speaking 4–5-year-olds and found bidirectional relations between vocabulary and ToM as well as a unidirectional relation from sentence comprehension to ToM but no relation between sentential complementation understanding and ToM. This contrasts with work by de Villiers and colleagues (de Villiers & Pyers, 2002; de Villiers, 2007) which established a link between complex syntactic abilities such as sentential complementation and ToM. Cross-linguistic findings appear to favour a general language hypothesis in ToM development over the specific role of any individual language competency (c.f. De Mulder et al., 2019; but see also Ebert, 2020).

From birth, newborns are oriented to socially and communicative relevant information: They distinguish direct from averted gaze, prefer infant-directed speech, and can differentiate positive and negative affective vocalisations (c.f. Falkum, 2019). Applying early social referencing skills, even 1- and 2-year-olds take others' internal states (first emotion, then knowledge) into account in drawing conclusions about consequences for themselves (Veneziano, 2009). As children acquire greater understanding of different communicative functions, they gain access to informative uses of language for the purpose of persuasion (Veneziano, 2009). At 3 years old, children begin using cognitive terms (e.g. 'know', 'think', 'imagine', 'guess', 'believe', 'understand') to talk about thoughts, beliefs and imagination (Grazzani & Ornaghi, 2012). Later, comprehension of mental state language continues to play a significant role in explaining children's performance on ToM tasks at primary school age and even into middle childhood (Grazzani & Ornaghi, 2012). Work with children between the ages of 5–12 has also shown that language skills mediate the relation between age and cognitive as well as affective ToM but in different ways at different ages (Bigelow et al., 2021). Research has also indicated that adult human mentalising abilities are typically limited to fifth-order intentionality (e.g. "I think that you believe that I suppose that you want me to imagine that...") which may initially cognitively scaffold recursive syntax abilities (Oesch & Dunbar, 2017). In turn, recursive syntax abilities then appear to provide a linguistic schema which facilitates conceptualisation of higher order intentionality within a positive feedback system (Oesch & Dunbar, 2017). Over the course of development, language (including gesture) and ToM may thus alternate in supporting progress within each other's domains.

4 Metaphor Comprehension

Processing ambiguous and non-literal metaphors requires a listener to make mappings between concepts, contexts, and linguistic forms in order to access a speaker's intended meaning. Children's development and mastery of these skills occurs over an extended period, continuing into adolescence and young adulthood (for more detailed overviews of metaphor skill development across the TD, LT, and DLD populations, please see Crawshaw et al., 2025a, 2025b; Appendices A and B). Metaphor comprehension and generation may also utilise different cognitive resources (Kasirer & Mashal, 2017).

Metaphors may draw on mental, perceptual, spatial, and aspectual imagery as well as language, inference-making, and ToM skills to make a comparison or an analogy and to reach accurate conclusions about it. Metaphors also possess referential and representational properties, analogous to gesture. Prior work has proposed that the relational structure in human thinking is important for each of these skills (gesture, language, ToM, metaphor) and has already drawn parallels between the attributive and relational qualities of metaphor and gesture (Cooperrider & Goldin-Meadow, 2017). Cooperrider and Goldin-Meadow (2017) observed that both metaphors and metaphoric gestures (linguistic and visual counterparts) can focus on attributes or relations, labelling the relational metaphoric gestures as 'analogical'. The authors identified that these 'analogical' gestures represented different types of relations as well as different degrees of relational complexity, fitting together within greater analogical models (Cooperrider & Goldin-Meadow, 2017, p.719). These ideas are highly relevant to and consistent with the aforementioned works (discussed in section 2.3) by Lakoff and Johnson (1980) on the metaphorical conceptual system underlying human communication and by McNeill and Levy (1982) indicating that gesture visually conforms to these common families of linguistic metaphor. It appears that the processing of gestures and metaphors utilise very similar underlying cognitive processes. Interesting links between them have already been documented in existing neuroscientific work, such as their common recruitment of general semantic processing areas in the brain (Joue et al., 2020; see also Ibáñez et al., 2010; for a discussion of these as well as a detailed consideration of the relations between gesture, language, ToM, and metaphor, see Appendix A: Crawshaw et al., 2025b). Moreover, analogical perception (identifying similarities across objects) and handling alternative naming (accepting that the same referent might have two labels) have been characterised as important for metaphor comprehension (Di Paola et al., 2020; Pouscoulous & Tomasello, 2020). Working with 3- and 4-year-olds, Di Paola et al. (2020) also found that children with both better alternative naming and analogical skills performed better on a metaphor comprehension task.

Processing iconic gestures arguably also taps analogical perception, i.e. recognising a gesture's similarities to an intended referent, and perhaps even alternative naming to some extent, given that a referent can be labelled both verbally and visually. Even pointing gestures could be viewed as an early form of relational thinking and communication, since they establish a spatial (or metaphorically temporal, c.f. Cienki & Müller, 2008) relation between the interlocutors and a target referent (or a past or future context, c.f. Cienki & Müller, 2008). Cienki and Müller (2008, p.486–487) described metaphoric gestures as those which had the potential to engage active cross-domain mappings to express thoughts or emotions, engaging in a cognitive process of perceiving something through the lens of something else. This is in essence relational thinking. Similar to existing findings on wider iconic gesture, Cienki and Müller (2008) additionally observed that metaphoric gestures and speech can relate to one another in different ways: (1) Speech and gesture may both express the same metaphor, (2) gesture may express a metaphor while the co-occurring speech does not, (3) speech and gesture may concurrently express different metaphors, (4) gesture may express a metaphor which is not used in that language, and (5) metaphoric gestures can vary or be consistent across cultures. These observations further emphasise that metaphorical thinking has a broader cognitive basis and metaphorical communication is not limited to speech alone.

It is also worthy of note that ToM has been shown to be sensitive to analogical skills and relational processing (Hoyos et al., 2020) as well as alternative or ‘dual’ naming (c.f. Doherty & Perner, 2020). A (2020) study by Hoyos et al. demonstrated that training 4.5–5-year-olds on analogical comparison and abstraction improved their performance on ToM false belief tests. The authors highlighted that analogical processing involves identifying common relational structure and patterns (including causal and functional properties), supporting abstraction and acquisition of more generalised schema necessary for ToM, e.g. person A believes performing action B will achieve scenario C, therefore person A is likely to take action B to realise scenario C (Hoyos et al., 2020). Deschrijver and Palmer (2020) also presented a notable reframing of ToM as a relational rather than a representational form of mentalising, whereby it was proposed that we relate others’ mental states to our own instead of attributing content to them. Falkum and Köder (2020) drew interesting parallels between the cognitive processes involved in overextension, pretence (also a precursor to ToM), and metaphor, suggesting that overgeneralisations on the basis of perceptual similarity potentially pave the way for early metaphor production. Overextension or overgeneralisation are arguably also forms of relational thinking. Findings linking analogical thinking, gesture, metaphor, and ToM, in addition to indications that they conform within wider models are intriguing. It might follow that early relational thinking supports gesture and language acquisition, i.e. responding to and then understanding the intentionality behind a pointing gesture or recognising that a certain combination of sounds relates to a specific object or concept. In turn, gesture and language, which are fundamentally referential and representational, might then facilitate more complex developments in relational or analogical thinking necessary for ToM and metaphor comprehension within a positive feedback system.

Metaphor processing and gesture have already been linked and considered within the framework of embodied cognition (c.f. Khatin-Zadeh, 2023; Khatin-Zadeh, Hu, et al., 2023; for further discussion of gesture and metaphor in relation to embodied and embedded cognition, see also Appendix A: Crawshaw et al., 2025b). However, theories linking metaphor comprehension to relational thinking and those connecting it to embodied cognition are not mutually exclusive. Embodiment and, by extension, embeddedness are arguably both also forms of (or at the very least draw upon) relational thinking. For example, using our own bodily properties to inform our perception and conceptualisation of our environment, simulating perceptual and motor properties to make connections between words and real-world referents, relating ourselves to external contexts by embedding internal representations of ourselves within them and modelling causal relations, etc. An understanding of internal self as opposed to external other is both representational and relational. Using "…grounded simulations of linguistic experience [to] help us acquire, represent and use new concepts" (Dove et al., 2020, p.2451) also draws on relational cognitive processes. With reference to Lakoff and Johnson's (c.f. 1980; previously discussed here in section 2.3) conceptual metaphor theory, Pecher et al. (2011) argued that situations (i.e. contexts) are needed to facilitate full representation of meaning in abstract concept processing, grounded in embodied sensory-motor image schemas. Prior to Pecher et al. (2011), Barsalou and Wiemer-Hastings (2005) emphasised the importance of situational content and focus (analogous to relational content and focus), theorising that conceptual representations stem from perceptual experiences which occur within wider perceived situations. Barsalou and Wiemer-Hastings (2005) also reasoned that we first need direct experience of an abstract concept in order to provide structure for a concrete metaphorical mapping. An experience, situation or context inherently establishes a relation, thus relational thinking may support embodied cognition. In order to create and represent categories, linguistically and cognitively, or make a metaphorical mapping between a concrete and abstract concept (c.f. Lakoff & Johnson, 1980), we need to perceive a relation or lack thereof between two or more items. Association (c.f. Barsalou & Wiemer-Hastings, 2005), generalisation and abstraction (c.f. Dove et al., 2020) are thus essentially relational. At a higher level, metalinguistic awareness and metacognition (relevant to metaphor comprehension and ToM) could similarly be considered as conscious utilisations of relational cognitive processes, understanding something (or someone else) through the lens of something else (or one's own self).

In summary, metaphor comprehension is a complex skill, developing across an extended time-period and depends upon a number of related abilities. Metaphor processing has been linked to wider theories about analogical or relational perception and thinking as well as to theories about embodied cognition and representation, however these do not appear to be mutually exclusive. Metaphors can be expressed in speech but also through gesture and processing them can draw upon wider language and ToM skills. Across a range of prior works, metaphor comprehension, gesture, spoken language, and ToM have separately all been linked to each other developmentally. They have also each been connected with relational thinking or embodied cognition paradigms. This indicates that these skills are all related and that they may tap common, underlying cognitive processes. Bearing these strong associations in mind, the following chapters will turn to a presentation and overview of the over-arching research questions and publications supporting this dissertation (Chapter 5: *Overarching Research Question and Overview of Supporting Publications*) as well as an overall discussion of these publications and their relevance to one another (Chapter 6: *Overall Discussion*), before final conclusions are addressed (Chapter 7: *Conclusions and Future Directions*).

5 Overarching Research Question and Overview of Supporting Publications

Table 1

Overview of the two publications supporting the doctoral dissertation.

Study	**Publication I (Appendix A):** Crawshaw et al. (2025b)	**Publication II (Appendix B):** Crawshaw et al. (2025a)
Title	*Does Early Gesture Usage Contribute Alongside Oral Language to Later Theory of Mind Performance and Metaphor Comprehension? Indications from a Longitudinal Study with Children Aged 1–9 Years*	*Late Talkers' Language, Metaphor, Theory of Mind, and Reading Skills at 9 Years*
Authors	Camilla E. Crawshaw, Carina Lüke, & Ute Ritterfeld	Camilla E. Crawshaw, Carina Lüke, & Ute Ritterfeld
Journal	Published in: *Language Acquisition*, 1–31.	Published in: *Journal of Speech, Language and Hearing Research (JSLHR)*, *68*(3), 1038–1055.
Research Question(s)	Can early gestural behaviour (pointing at age 1;0 both in terms of hand shape and communicative motive; iconic gesture comprehension at age 3;0) predict outcomes at age 9;0 in language (productive vocabulary, productive and receptive grammar), ToM (hidden emotions, sarcasm, explicit false belief; Happé's Strange Stories), and metaphor comprehension skills?	(1) Are language and reading abilities impacted in (German-speaking) LTs at age 9;0? (2) Are metacognitive and metalinguistic (ToM and metaphor comprehension) abilities affected in 9-year-old children with a history of being a LT?
Methodological Approach	Pearson's correlational analyses Stepwise multiple linear regression analyses Mediation analyses (Hayes, 2018) Effect sizes and post-hoc statistical power were analysed. The Bonferroni-type Benjamini and Hochberg (1995) method was also applied to control the false discovery rate when conducting multiple comparisons.	Non-parametric group comparisons using Mann-Whitney-U tests. Effect sizes and post-hoc statistical power were analysed. The Bonferroni-type Benjamini and Hochberg (1995) method was also applied to control the false discovery rate when conducting multiple comparisons.

Two research gaps were addressed within the following publications (Crawshaw et al., 2025a, 2025b; Appendices A and B) supporting this doctoral dissertation. Firstly, to this author's knowledge, no previous study had holistically considered gesture, spoken language, ToM, and metaphor comprehension in combination with one another and within the context of a longitudinal study. Secondly, while prior work has addressed gesture, spoken language, ToM, and metaphor comprehension within the TD and DLD populations, it appeared that no study had yet explored ToM and metaphor comprehension within the 'recovered' LT population. The two supporting publications aimed to contribute towards closing the above-mentioned research gaps by posing the following specific research questions. Publication I (Appendix A: Crawshaw et al., 2025b): Can early gestural behaviour (pointing at age 1;0 both in terms of hand shape and communicative motive; iconic gesture comprehension at age 3;0) predict outcomes at age 9;0 in language (productive vocabulary, productive and receptive grammar), ToM (hidden emotions, sarcasm, explicit false belief; Happé's Strange Stories), and metaphor comprehension skills? In Publication II (Appendix B: Crawshaw et al., 2025a) two questions were posed: (1) Are language (productive vocabulary, productive grammar, receptive grammar) and reading abilities impacted in (German-speaking) LTs at age 9? (2) Are metacognitive and metalinguistic (ToM and metaphor comprehension) abilities affected in 9-year-old children with a history of being an LT?

Previous work has uncovered deep connections across each of these competencies (gesture, language, ToM, and metaphor comprehension), albeit taking more of a one-to-one approach. This doctoral dissertation therefore seeks to incorporate all of these language-related skills within a wider, more holistic perspective as well as consider them with respect to the LT context, a population which has historically experienced challenges in language acquisition. The overarching research question of this dissertation is thus: *How might gesture, language, ToM, and metaphor comprehension be connected in development?* The publications supporting this dissertation address this question from two different angles. Publication I (Appendix A: Crawshaw et al., 2025b) dives more deeply into each competency as both individual and related constructs, reflecting upon their theoretical backgrounds and potential interrelations, before addressing them from a predictive, longitudinal perspective spanning infancy to middle childhood. Publication II (Appendix B: Crawshaw et al., 2025a) takes a more cross-sectional approach to consider how these competencies may or may not look differently in a group of former LTs compared to their TD peers at age 9;0, reviewing and comparing the existing perspectives from prior work regarding each competency within the TD and DLD populations. An overview of each supporting publication is presented in Table 1. Their specific details and main results are summarised in the following text.

Publication I. Existing work had already demonstrated that: (a) language ability can be predicted and mediated by earlier gestural skills (e.g. Colonnesi et al., 2010; Lüke et al., 2020; Mainela-Arnold et al., 2014; Wray et al., 2016). It had also identified that (b) later ToM skills could be predicted by various language abilities (Durrleman, 2020; Durrleman & Delage, 2020; Milligan et al., 2007; Osterhaus & Koerber, 2021b) and that both are related to metaphor comprehension skills (for a discussion, see Pronina et al., 2023). However, prior research had not yet clearly established (c) a direct relationship between ToM development and early deictic or iconic gesture use (Carlson et al., 2005; Cochet et al., 2017; Colonnesi et al., 2008; Sodian & Kristen-Antonow, 2015). Nor had it yet determined d) a clear-cut relationship between metaphor comprehension skills and early gestural communicative competencies, although prior theory and evidence had already indicated some parallels between the representational and intentional natures of metaphor and gesture (Camaioni et al., 2004; Doherty, 2000; Doherty & Perner, 1998, 2020; Khatin-Zadeh, 2023; Liszkowski & Tomasello, 2011; McNeill, 2014). Prior work had also revealed that a point’s handshape (i.e. using the index-finger rather than the whole hand) is more important than its communicative motive in predicting spoken language ability (Lüke, Grimminger, et al., 2017). Given the separately documented links between each of gesture, language, ToM, and metaphor comprehension in existing work, the work in this publication was aimed at bringing these competencies together within a longitudinal perspective to explore common processes which may underlie them. Another important aim was to expand focus on the specific role of the index finger in pointing and clarify how this might be relevant for wider language-related skills such as ToM and metaphor comprehension. To this end, the following research question was posed: *Can early gestural behaviour (pointing at age 1;0 both in terms of hand shape and communicative motive; iconic gesture comprehension at age 3;0) predict outcomes at age 9;0 in language (productive vocabulary, productive and receptive grammar), ToM (hidden emotions, sarcasm, explicit false belief; Happé's Strange Stories), and metaphor comprehension skills?*

To test the research question and establish whether early gestural development (index-finger pointing at age 1;0) could be identified as a predictor of outcomes in language, ToM, and metaphor comprehension skills at age 9;0 (alongside the mediator variables iconic gesture comprehension or language skills at age 3;0), further data collection was carried out with *N* = 35 children who had already participated within a longitudinal study between the ages of 1;0–6;0 (c.f. Lüke, 2015; Lüke et al., 2020, 2019; Lüke, Grimminger, et al., 2017; Lüke, Ritterfeld, et al., 2017; Rohlfing et al., 2022). A sequence of Pearson's correlational, stepwise regression and mediation analyses were conducted on the resulting data. Post hoc effect size calculations and power analyses were also carried out, and corrections were applied to control for the false discovery rate when conducting multiple comparisons. Results from the Pearson's correlation analyses revealed a statistically significant correlation between index-finger pointing at 1;0 and only productive vocabulary and metaphor comprehension at age 9;0. The various communicative motives for pointing at age 1;0 were not correlated with any outcome variables at age 9;0 and were excluded from further analysis. Iconic gesture comprehension at age 3;0 was significantly correlated with all outcome variables at age 9;0 except for the ToM Strange Stories task. The composite value of language skills at 3;0 was significantly correlated with all outcome variables at age 9;0. Results from the stepwise regression models showed that index-finger pointing at 1;0 could directly predict metaphor comprehension and vocabulary skills at age 9;0. Results from the mediation analyses demonstrated that index-finger pointing at 1;0 could also indirectly predict metaphor comprehension, performance on one ToM task (ToM Scale), productive vocabulary skills, and receptive grammar skills at 9;0 via both mediation pathways at 3;0 (iconic gesture comprehension, composite language skills). In addition, it could indirectly predict performance on the second ToM task (Strange Stories) at 9;0 via only the language skills at 3;0 mediation pathway.

Publication II. Prior work (e.g. Rescorla, 2002, 2009; Rescorla et al., 1997; Stothard et al., 1998) had identified differences in LTs (both with and without a later diagnosis of DLD) from their TD peers across a range of language outcomes, continuing into adolescence. Findings regarding reading-related skills were contradictory, indicating that they might be differentially impacted across developmental stages (c.f. Bishop & Edmundson, 1987; Rescorla, 2009; Stothard et al., 1998). Previous research had suggested that children with DLD experience delays in both ToM and metaphor comprehension (e.g. Nilsson & de Lopez, 2016; Norbury, 2005; Bühler et al., 2018), but no study had yet investigated whether LTs who appear to 'recover' also face challenges with these skills. Other work had separately shown that later performance on both ToM (e.g. Milligan et al., 2007) and metaphor processing tasks (e.g. Deckert et al., 2019; Kalandadze et al., 2018) can be predicted by earlier language ability. Metaphor comprehension and ToM skills had also been associated with one another (e.g. Lecce et al., 2019; Norbury, 2005). Language ability had already been conceptualised as a spectrum ("language endowment spectrum"; c.f. Ellis Weismer, 2007, p. 84), and it was hypothesised in Publication II (Appendix B: Crawshaw et al., 2025a) that LTs would thus likely fall somewhere between children who were TD and those with DLD along spectra or continua of language-associated skills like ToM and metaphor comprehension. In response to existing work, the following two research questions were posed: *(1) Are language (productive vocabulary, productive grammar, receptive grammar) and reading abilities impacted in (German-speaking) LTs at age 9? (2) Are metacognitive and* *metalinguistic (ToM and metaphor comprehension) abilities affected in 9-year-old children with a history of being an LT?*

To test these research questions, the *N* = 35 children who had participated in the study reported in Publication I, were reanalysed and compared in terms of their historical status within the wider longitudinal study at age 2;0 as either TD (*n* = 27) or LT (*n* = 8). Nonparametric group comparisons were conducted using Mann-Whitney *U* tests as a result of the small and uneven group sizes. Post hoc effect size calculations and power analyses were performed in addition to the application of corrections to control for the false discovery rate when conducting multiple comparisons. Results from the Mann-Whitney *U* tests showed that the LTs performed significantly less well than their TD peers across the measures of productive vocabulary, receptive grammar, metaphor comprehension, and in the Strange Stories task but not on the ToM Scale tasks or the reading measure. There was a trend towards a slightly lower performance on the productive grammar measure by the LTs but this did not reach statistical significance. After applying corrections to control for the false discovery rate when conducting multiple comparisons, significant group differences were no longer found on the productive vocabulary measure.

The above studies will be discussed in relation to the overarching research question in the following chapter (Chapter 6: *Overall Discussion*). The relations between skills and potential common underlying processes will be discussed in section 6.1, the idea of a developmental continuum of language-related skills will be discussed in section 6.2, and the limitations of these studies will be addressed in section 6.3.

6 Overall Discussion

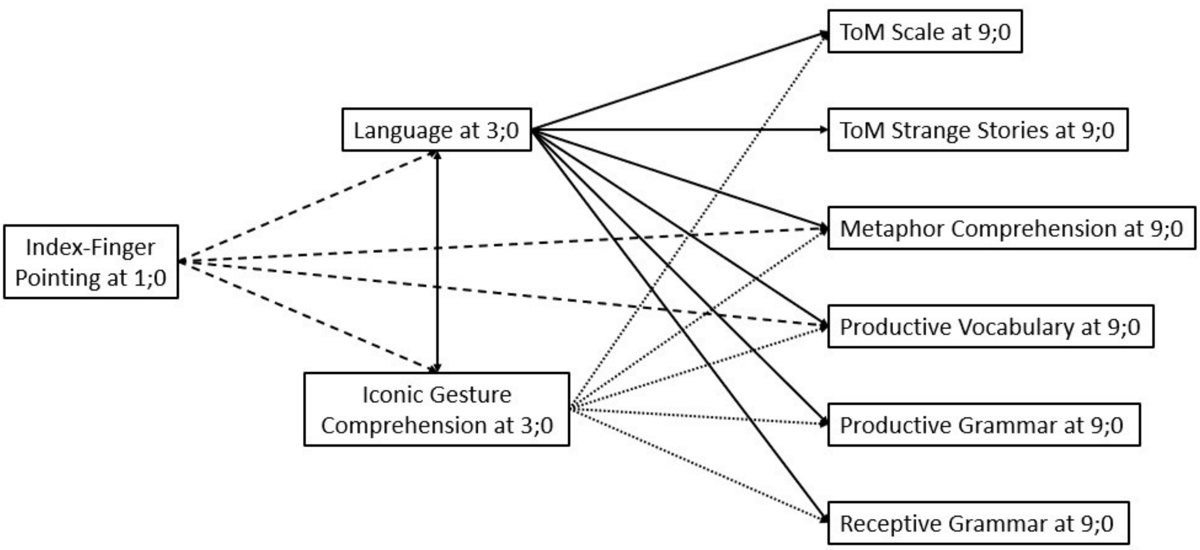
Prior work has taken great interest in language acquisition, gestural development and ToM, while metaphor comprehension has received slightly less focus. However, these constructs have largely been considered either in isolation or at most in relation to one or two of the others. Research with atypically developing populations has also not yet addressed these constructs equally or from a holistic perspective, accounting for a wider network of associated skills. This is surprising given the great extent of common underlying processes supporting each of these competencies. Through highlighting broader connections between gesture, language, ToM, and metaphor comprehension, as well as situating them within a more holistic developmental framework, this doctoral dissertation seeks to address these research gaps as well as to present theoretical and practical insights to support future work in these areas. The following text will discuss the results of the two supporting publications (Crawshaw et al., 2025a, 2025b; Appendices A and B) in relation to one another and in response to the overarching research question: *How might gesture, language, ToM, and metaphor comprehension be connected in development?* Following a reflection on the limitations of this work in section 6.3, its potential implications for future research will be presented in Chapter 7: *Conclusions and Future Directions*.

6.1 Relations between Skills and Common Underlying Processes

To establish how gesture, language, ToM, and metaphor comprehension might be connected across the developmental span, one must first consider their individual relations as well as how they might similarly or differentially utilise common underlying processes. In Publication I (Appendix A: Crawshaw et al., 2025b), longitudinal and cross-sectional data regarding participants' performances in each of these areas were analysed using Pearson's correlational (see Figures 1 and 2), stepwise regression, and mediation analyses. To address this first point, the relations between each competency identified by those analyses will be discussed briefly below and linked to wider theoretical paradigms (the findings are already discussed in considerable detail within the *Discussion* section of Publication I, Appendix A: Crawshaw et al., 2025b).

Figure 1

Visual representation of statistically significant Pearson's correlations between variables longitudinally, after applying corrections to control for the false discovery rate when conducting multiple comparisons (data reported in Appendix A: Crawshaw et al., 2025b).



The longitudinal findings regarding gesture will be considered first, since this constitutes the data's chronological starting point. As can be seen from Figure 1 above, index-finger pointing at 1;0 correlated significantly with both iconic gesture comprehension and language skills at 3;0 in addition to metaphor comprehension and productive vocabulary at 9;0, measured eight years later. Iconic gesture comprehension correlated significantly with language skills at 3;0 in addition to all language skills (productive vocabulary, productive grammar, receptive grammar), metaphor comprehension, and the ToM Scale (but not the ToM Strange Stories task) at age 9;0; the composite language skills at 3;0 correlated with all outcome variables at age 9;0. Analysis of the data also produced statistically significant stepwise regression models demonstrating that index-finger pointing at 1;0 could directly predict metaphor comprehension and vocabulary skills at 9;0. In mediation models with iconic gesture comprehension at 3;0 as the mediator, index-finger pointing at 1;0 could also indirectly predict productive vocabulary skills, receptive grammar skills, metaphor comprehension, and performance on the ToM Scale at 9;0 (via the mediation pathway of composite language skills at 3;0, it could indirectly predict the same skills in addition to performance on the ToM Strange Stories task). The findings from Publication I (Appendix A: Crawshaw et al., 2025b) thus indicate that both early gestural production and comprehension (index-finger pointing and iconic gesture comprehension) are strongly linked to later metaphor comprehension. These skills were measured many years apart and the extent to which they are associated suggests that they may utilise similar underlying cognitive processes, e.g. analogical perception as well as wider semantic, conceptual, visuospatial, representational skills (these are extensively discussed in Publication I, Appendix A: Crawshaw et al., 2025b; see also Chapter 4). If gestures are manifested representational substitutes standing in analogically for various ideas, objects, actions or relations (Novack & Goldin-Meadow, 2017, p.652), then it makes sense they would share commonalities with metaphors, which are verbal expressions of analogical relations. The findings further demonstrated that early gestural production could be directly linked to productive vocabulary skills (one specific aspect of language) eight years later, indicating that these skills may also share some sort of facilitative relationship. At a basic level, the physical act of deictic reference might spontaneously initiate lexical access and retrieval. As already discussed in Publication I (Appendix A: Crawshaw et al., 2025b, p.23), the task used to assess the children's productive vocabulary (P-ITPA; Esser et al., 2010) is actually both lexical and semantic. Index-finger pointing at 1;0 may thus be associated with metaphor comprehension and productive vocabulary skills in different ways, since semantic knowledge has been identified as a key component in metaphor comprehension (Deckert et al., 2019; Vulchanova et al., 2019). This may partially explain a link to iconic gesture comprehension, for which semantic skills are also very important. Semantic skills may potentially extend beyond lexical skills in drawing on wider associative skills, analogical perception and relational thought processes.

A further indirect link between gestural production and later receptive grammar skills was found via the mediator of gestural comprehension at 3;0, while no links were found in relation to productive grammar at 9;0. The mixed association between gesture and grammar is interesting. Prior work has discussed how gesture can function as a physical manifestation and realisation of grammatical processes including modal stances, aspectual contrasts, deictic reference, and causality, with integration on the syntactic, functional, and conceptual levels (Lapaire, 2011). The schematicity of gestural expression and grammatical meanings also correlate with one another (Lapaire, 2011). The inherent dynamicity of gestural imagery is often reflected verbally in grammatical morphemes such as the present progressive tense, for example: “She was walk[*ing*] past me…”, co-occurring with a gesture where the fingers visually walk from one side of the gesture space to the other. In an earlier stage of work with the same sample analysed in Publication I (Appendix A: Crawshaw et al., 2025b), index-finger pointing at 1;0 could additionally predict productive grammar at ages 5;0 and 6;0 both in similarly conducted stepwise regression models and in mediation models via the same iconic gesture comprehension at 3;0 mediator pathway (Lüke et al., 2020). It is possible that these processes have largely become automatised by the age of 9;0 such that previous effects are no longer visible. Alternatively, the small sample size and consequent lower statistical power might have masked smaller effects; this should be followed up in a future study with a greater number of participants. Conversely, we may actually see the significant correlation between gesture comprehension and receptive grammar (as well as the successful mediation effect) for precisely the same reasons discussed above. Since both are receptive skills, it is possible that similar, ‘online’, embodied mental simulation skills are tapped when processing dynamic visuospatial gestural imagery and linguistic expressions of grammar (c.f. Gesture as simulated action framework: Hostetter & Alibali, 2019, 2008).

Gestural production was indirectly linked through the mediator gestural comprehension to only one of the two ToM tasks (ToM Scale) at age 9;0, while the mediator composite language skills linked it to both ToM tasks (ToM Scale and ToM Strange Stories).

Clearly early gestural production and comprehension as well as language skills are still important for ongoing ToM development (this is in line with extensive prior work, e.g. Astington & Baird, 2005; Doherty, 2008; Durrleman, 2020; Durrleman & Delage, 2020; Ebert, 2020; Milligan et al., 2007; Osterhaus & Koerber, 2021b). As already discussed in section 2.1, gestures can visuospatially embed and portray different perspectives or viewpoints (e.g. first-person, third-person, dual) while spoken language encodes them linguistically (Demir et al., 2015; McNeill & Levy, 1982; Parrill et al., 2018; Tomasello, 2019). The lack of direct effects however may reflect a certain degree of abstraction from the early functions of deictic gesture production and beginnings of ToM in contrast to later more sophisticated ToM processes. Parallels could be drawn between this and Fenici's (2012) discussion of early embodied ToM processes in contrast to later embedded ToM processes, reflecting changes in ToM operation over time. The picture from experimental work suggests that more general pointing (e.g. initiating joint attention) is predictive of early ToM development (Brandone & Stout, 2023; Cochet et al., 2017; Colonnesi et al., 2008; Sodian & Kristen-Antonow, 2015). Later, more sophisticated ToM development may then be scaffolded and supported by language (integrated with co-speech gesture) as an external structure (c.f. Fenici, 2012), incorporating a network of other associated cognitive skills and processes (this is further discussed in relation to the findings of Publication I in Appendix A: Crawshaw et al., 2025b).

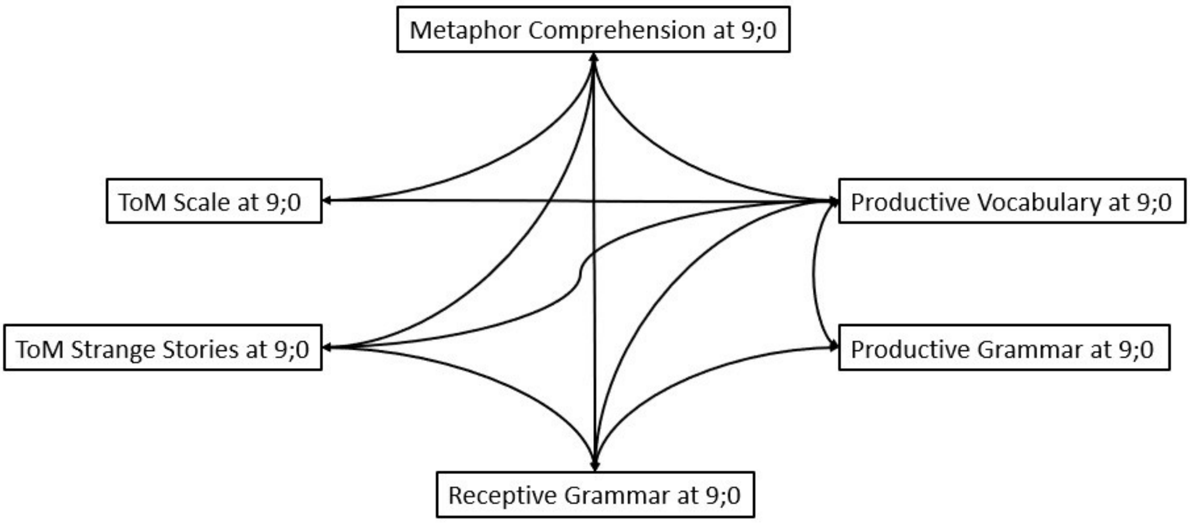
In this vein, existing work has documented changes in ToM ability across middle childhood, although these studies are much fewer in number than accounts of early childhood (Apperly, 2021; Devine & Lecce, 2021). ToM development follows an extended trajectory into middle childhood and adolescence, rooted in experience of prior life events, as children gain an increasingly more sophisticated awareness of the potential interactions between different mental states (Devine & Lecce, 2021; Lagattuta & Kramer, 2021). Executive function (e.g. working memory) also contributes to ToM development: Children's executive control improves as they develop which supports further improvements in both their ToM capacity and language abilities (Apperly, 2021; Lagattuta & Kramer, 2021).

Tomasello (2019, p.35) observed that individual differences in ToM may emerge from a transactional causal model of maturation, whereby maturational capacities are only stimulated through individuals' transactions with their (social) environments. ToM is thus a diverse construct encompassing multiple aspects of mental state reasoning which vary across developmental timetables: Both age as well as social context may interact in predicting when and how the ToM network is engaged (Lagattuta & Kramer, 2021, p.28; Warnell & Redcay, 2021, p.84). Given the huge scope of ToM as an umbrella construct and the significant role of individual social experiences in its development, it is extremely difficult to pinpoint and precisely measure the contributions of specific facilitative factors. Since the samples analysed in Publications I and II (Crawshaw et al., 2025a, 2025b; Appendices A and B) are relatively small, individual differences could have skewed the data. It is impossible to control for every past social experience a child has ever had and its individual contribution to their personal ToM development so one can only try to explore common supportive processes at the more generalised level. This may go some way toward explaining the historically variable findings regarding ToM and gesture. The early insights presented in this dissertation and its supporting publications should also be built upon and addressed within larger samples in future, more targeted work.

Since ToM may cover too broad a network of associated mentalising skills, it may also be helpful to break ToM task batteries down into smaller components or group them on the basis of specific commonalities before relating them to other developmental constructs. This was one rationale for intentionally not compounding the ToM Scale and ToM Strange Stories within Publications I and II (Crawshaw et al., 2025a, 2025b; Appendices A and B). Neuroimaging work has demonstrated that performance on different types of ToM tasks can also show very little and even no relation to one another for a host of reasons including limitations regarding insufficient age ranges, sample sizes, missing data, and a lack of sensitivity in measurement technique (c.f. Warnell & Redcay, 2021; specific limitations regarding the ToM Scale task will be addressed in more detail within section 6.3). As can be seen from Figure 2, performances across the two ToM tasks in the study supporting this dissertation were not correlated with each other. Potential differences in task demands as well as the idea that these may fall under different subareas of ToM are already discussed at length in Publication I (Appendix A: Crawshaw et al., 2025b). It may also be interesting to consider cross-sectional correlations across the areas at age 9;0 (Figure 2) to further explain the different mediation effects regarding gesture, language, and ToM skills, as well as to illuminate potential groupings and shared pathways in underlying processes.

Figure 2

Visual representation of statistically significant Pearson's correlations between variables cross-sectionally at age 9;0, after applying corrections to control for the false discovery rate when conducting multiple comparisons (data reported in Appendix A: Crawshaw et al., 2025b).



While they did not correlate with one another, both ToM tasks were significantly correlated with metaphor comprehension and productive vocabulary at age 9;0, while the ToM Strange Stories task was also significantly correlated with receptive grammar (see Figure 2). Metaphor comprehension and productive vocabulary were both correlated significantly with one another as well as with receptive grammar at 9;0. Productive grammar was only significantly correlated with receptive grammar and productive vocabulary at 9;0. The visual representation of these cross-sectional correlations at 9;0 (Figure 2) implies the potential existence of different, intricate underlying subnetworks of these associated skills, where specific shared pathways may be accessed in different ways and at different times in supporting more complex operations.

One potential subnetwork indicated by the correlations presented in Figure 2 may constitute language skills comprising productive and receptive grammar, productive vocabulary, and likely also including receptive vocabulary. A language subnetwork could further incorporate gestural production and comprehension, since together they form an integrated system (c.f. Chapter 2). Receptive vocabulary was unfortunately not assessed in the studies reported in Publications I and II (Crawshaw et al., 2025a, 2025b; Appendices A and B). It would likely form part of a language subnetwork and probably be correlated with each of the other language skills assessed. However, it would also be interesting to consider it in relation to the other skills addressed by the study. Since it was not included, one can only look to existing prior work for an idea of how it might be integrated within wider associated skill networks. Relevant to the longitudinal skill associations, Doğan et al. (2024) conducted a recent study with 92 children aged 1;10–2;6 years and found that receptive vocabulary and iconic gesture comprehension were positively related in children until the age of 2;2. This finding suggests that the iconic gesture comprehension measure conducted at age 3;0 in Publication I (Appendix A: Crawshaw et al., 2025b) would likely not have been linked to a later receptive vocabulary measure at age 9;0 (had it been included) but this hypothesis should be confirmed by future work. In terms of metaphor comprehension, Huang et al. (2015) carried out a study with children aged 7.4–12.5 years and found that receptive vocabulary performance correlated significantly with metaphor comprehension but not with measures of irony and sarcasm. These findings suggest that a link between metaphor comprehension and receptive vocabulary might have been found in Publication I (Appendix A: Crawshaw et al., 2025b) had this measure been included. However, sarcasm also constituted one of the ToM Scale subtests used in the studies reported in Publications I and II (Crawshaw et al., 2025a, 2025b; Appendices A and B) so it is possible that no relation would have been found between receptive vocabulary and the ToM Scale (but there are some further limitations regarding this measure which will be discussed in section 6.3).

On the whole, prior work does indicate that receptive vocabulary is relevant for ToM albeit less important than other language skills. In Milligan et al.'s (2007) meta-analysis, receptive vocabulary measures were found to have weaker relations to ToM skills than general language. The authors proposed this low effect size may result from receptive vocabulary tests being an isolated measure of language skills with little overlap with other abilities, requiring children to simply point to a target picture (Milligan et al., 2007, p.636). However ToM was only assessed in this meta-analysis in terms of performance on the false belief test (only one of the subtests comprising the ToM Scale measure employed in Publication 1, Appendix A: Crawshaw et al., 2025b). It is possible that receptive vocabulary may be more relevant for more advanced measures of ToM which also place higher linguistic demands, such as the ToM Strange Stories measure. In a longitudinal study, Ebert (2020) found that early receptive vocabulary (age 3;6) was significantly correlated with later ToM: both the false belief test at age 5;6 and Strange Stories at age 12;8. Receptive vocabulary at age 5;6 was significantly correlated with the ToM false belief test at age 5;6 and not the ToM Strange Stories at 12;8, while receptive vocabulary at age 12;8 was correlated with both ToM measures at 5;6 and 12;8 (Ebert, 2020). Ebert (2020) also found that while early receptive vocabulary was relevant for later ToM (Strange Stories), early receptive grammar was a stronger correlate. A measure of receptive vocabulary might thus be interesting to include in any future replication or extension of the studies reported in Publications I and II (Crawshaw et al., 2025a, 2025b; Appendices A and B).

The correlations in Figure 2 might suggest that the ToM tasks (which may constitute parts of a wider ToM construct) access a subnetwork of language (and gestural) skills differentially in both manner and timing. On a conceptual level, these links may reflect shared and underlying visuospatial, representational, analogical or relational thought processes, among others. On an operational level, ToM is also both understood and expressed through the medium of language, thus differential demands will influence performance. In terms of the specific language subskills, the ToM Scale appeared to be linked only with productive vocabulary, whereas the ToM Strange Stories task was linked with both productive vocabulary and receptive grammar, likely reflecting higher linguistic demands in understanding the longer sentences and texts (for a more detailed assessment of each task, please see Publication I, Appendix A: Crawshaw et al., 2025b). There is also some evidence for an emerging automaticity of certain ToM skills over the course of development (c.f. Schneider et al., 2017). It may be the case that the hidden emotions and explicit false belief subtests used in the study's ToM Scale measure proved simpler for the children and that they were beginning to be processed more automatically in contrast to the more complex ToM Strange Stories measure (Publication I, Appendix A: Crawshaw et al., 2025b), however this should be examined in future work. Given differential processing demands, subsequent work could also assess the impact of children's working memory (and how it may fit within a potential model of connected subnetworks) since it has been shown in past work to be correlated with recursive thinking skills (c.f. Valle et al., 2015).

Finally, metaphor comprehension appears to draw on processes connecting it to both potential subnetworks of ToM and language (including gestural) skills. Shared pathways may incorporate common and underlying visuospatial, representational, inferential, and analogical or relational thought processes. Metaphor comprehension and production would also probably constitute parts of a wider subnetwork of pragmatic skills, drawing similarly upon shared pathways with other subnetworks in contextually defined ways. As discussed at length in Publication I (Appendix A: Crawshaw et al., 2025b), metaphor comprehension has been strongly associated with language and particularly semantic skills (Deckert et al., 2019; Kalandadze et al., 2018; Vulchanova et al., 2019). Findings connecting metaphor comprehension and ToM skills have been mixed and it appears that the type of metaphor is the determining factor, i.e. whether or not it involves mental aspects (Lecce et al., 2017; Tonini et al., 2023; for a more detailed discussion, see Publication I, Appendix A: Crawshaw et al., 2025b). ToM may also differentially support metaphor comprehension depending on the specific task at hand; it may promote access to mental state terms or improve processing of context, supporting acquisition of more sophisticated pragmatic skills (Tonini et al., 2023). Prior work has further proposed that metaphor comprehension and other specific pragmatic skills (e.g. sarcasm or irony) fall at different points along a spectrum between ToM and structural language skills, within which those requiring integration of a speaker's intended meaning and mental state would fall closer to ToM ability (Pronina et al., 2023; also discussed in Publication I, Appendix A: Crawshaw et al., 2025b). However metaphor comprehension draws on wider underlying cognitive processes beyond structural language and ToM in addition to analogical perception and relational thinking. It may therefore be more helpful to conceptualise it within a holistic framework of interconnected subnetworks and shared pathways as proposed in this dissertation.

A framework of subnetworks connected by shared pathways of common underlying processes would fit within a number of theoretical paradigms. Possible links between gesture, language, ToM, and metaphor comprehension in terms of embodied and embedded cognition (e.g. Fenici, 2012; Goldin-Meadow, 1998; Hostetter & Alibali, 2019, 2008; McNeill & Levy, 1982; Khatin-Zadeh, 2023; Khatin-Zadeh, Farsani, et al., 2023; Khatin- Zadeh, Hu, et al., 2023; Pecher et al., 2011; see section 2.3) or in terms of relational thinking (e.g. Cienki & Müller, 2008; Cooperrider & Goldin-Meadow, 2017; Deschrijver and Palmer, 2020; Di Paola et al., 2020; Doherty & Perner, 2020; Falkum and Köder, 2020; Hoyos et al., 2020; Pouscoulous & Tomasello, 2020; see Chapter 4) are not mutually exclusive to each other or to the ideas behind shared intentionality (c.f. Tomasello, 2019; see Chapter 3). As discussed in Chapter 4, embodied or embedded cognitive processes could arguably be seen as forms of relational thinking. Using or simulating something to recognise, understand or refer to something else is ultimately utilising relational thought processes. Fundamentally, engaging in shared intentionality also requires an understanding of how internal 'self' may relate to external 'other'. It requires us to transfer our experiences of mental states to someone else's context, make connections to predict their potential behaviour and embed everything within a specific situational context.

In addition to linguistic and processing demands, variations in performance observed across ToM and metaphor comprehension depending on task and type could also reflect different extents to which they draw on embodied, embedded, relational, representational and attributive cognitive processes. Relevant here is work on metaphorical embodiment identifying that different metaphors engage sensorimotor systems to different degrees and that congruent gestural depictions prime comprehension of metaphors, equating to their embodied realisations (Khatin-Zadeh, 2023; Khatin-Zadeh, Farsani, et al., 2023; Khatin-Zadeh, Hu, et al., 2023). Findings suggesting that ToM and metaphor comprehension are more strongly linked when metaphors involve mental instead of physical aspects reveal further insights into the bigger picture (Lecce et al., 2019, Tonini et al., 2023). Metaphors are inherently relational but specific metaphors may differentially draw on additional embodied, embedded, and ToM processes. This would make sense, given that metaphors can express conceptual transfers of a wide range of different properties, qualities, actions, conditions, and states: physical, mental, visual, functional, perceptual, sensory, emotional, spatial, abstract, concrete, etc. Certain metaphors may be better communicated, processed and understood via particular means and their associated underlying cognitive pathways; gesture as a medium provides a dynamic visual and embodied complement to their linguistic encoding. ToM is another construct with complex multi-skill foundations and it may thus be possible that specific ToM tasks draw more heavily than others from certain associated skills via their underlying shared pathways. This may explain the findings of Publication I (Appendix A: Crawshaw et al., 2025b) which (a) link metaphor with a wide range of other skills and (b) reflect a disassociation between the two ToM tasks used which in turn showed both common and distinct links to the other skills under consideration.

6.2 Developmental Continuum of Language-Related Skills

If gesture, language, ToM, and metaphor comprehension really share common underlying processes, then it is of interest to consider these competencies all together within the developmental context of populations experiencing challenges across one or more of them. In doing so, we can perceive how these competencies take form when development is not proceeding as expected, which can provide us with fresh insights into both the existence as well as nature of their relations. Unlike children on the autistic spectrum or who have gone on to receive a diagnosis of DLD, former LTs have not received as much focus beyond structural language and reading skills. Earlier theories held that those children who appeared to catch up must have resumed a typical developmental trajectory but later work identified that former LTs actually continue to experience lower outcomes than their TD peers across language(-related) skills, even as they approach adulthood (c.f. Bates et al., 1995; Bishop & Edmundson, 1987; Horvath et al., 2022; Perry et al., 2023; Rescorla, 2009, 2011; Stothard et al., 1998). As children who may fall somewhere along a spectrum between those who are TD and those with DLD, considering the specific context of LTs could be particularly helpful in closing gaps between work addressing typical and atypical development. This section will aim to present another perspective on the '*in development'* component of the overarching research question of this dissertation (*How might gesture, language, ToM, and metaphor comprehension be connected in development?*).

In Publication II (Appendix B: Crawshaw et al., 2025a), the longitudinal sample of participating children reported on in Publication I (Appendix A: Crawshaw et al., 2025b) were analysed at age 9;0 from a cross-sectional perspective. The children who had a history of late talking were compared to their TD peers across language, reading, ToM, and metaphor comprehension tasks. Nonparametric group comparisons were conducted to analyse their relative performances. Former LTs performed worse than their TD peers on the productive vocabulary, receptive grammar, metaphor comprehension, and ToM Strange Stories tasks but not on the productive grammar, reading, or the ToM Scale tasks. Potential explanations for each of these outcomes are individually discussed in detail within Publication II (Appendix B: Crawshaw et al., 2025a) but it is also interesting to consider them from an even wider research perspective as well as how they relate to the findings of Publication I (Appendix A: Crawshaw et al., 2025b).

The following is a summary of potential explanations presented in Publication II (for more detail and relevant citations, see Appendix B: Crawshaw et al., 2025a): (1) Lower performance on the ToM Strange Stories and metaphor comprehension tasks may have resulted from (a) a related lower performance on the productive vocabulary and receptive grammar measures with early differences causing an enduring impact or (b) task demands and relevant verbal processing deficits. (2) The lack of difference in productive grammar performance may have resulted from (a) the LTs having caught up with their peers, (b) individual differences among the children or (c) as a consequence of task design. (3) The lack of difference in the reading measure may have been due to (a) crosslinguistic differences between this study and prior work (German is for example orthographically more transparent than English) or (b) again general differences in task design or demands. (4) The lack of difference on the ToM Scale was possibly due to its inherent limitations as a measure for assessing individual differences between children of the same age (this point will be addressed further in section 6.3: Limitations).

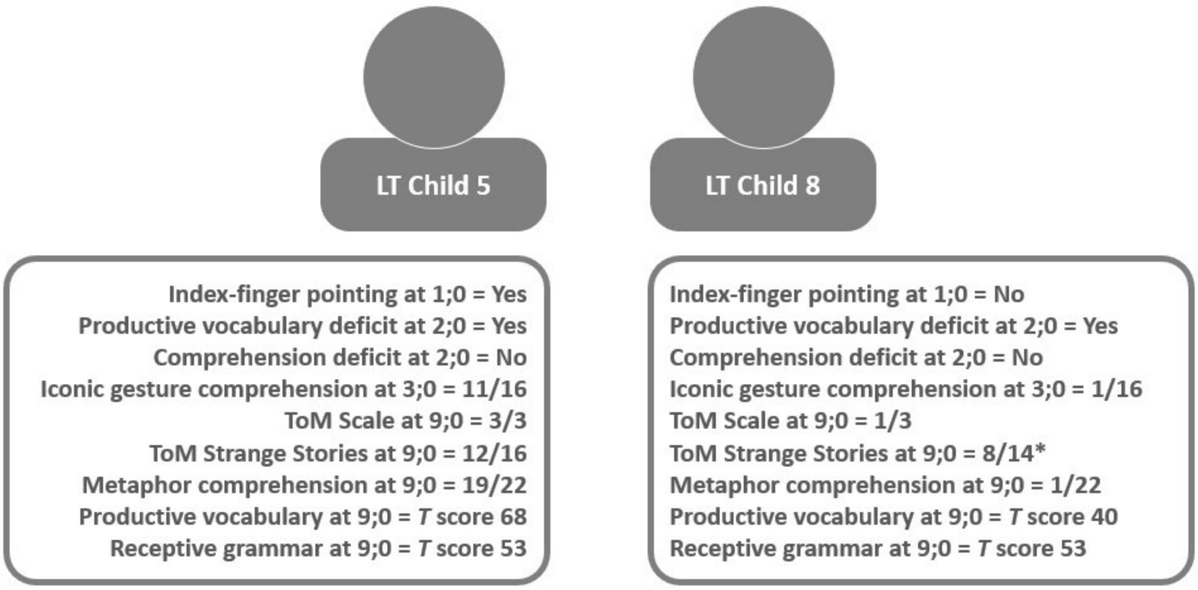
When these group findings and potential explanations are considered in relation to the longitudinal perspective of Publication I (Appendix A: Crawshaw et al., 2025b), some interesting parallels begin to emerge. Of course, caution in interpretation must be exercised since the analyses reported in both publications come from the same (small) sample; evaluation within the context of other existing work is thus essential. The first point concerns the LTs' lower performances across the productive vocabulary, receptive grammar, ToM Strange Stories, and metaphor comprehension tasks. Referring back to Figure 2 (presented in the previous section), it is striking that these four areas appear to form some kind of subnetwork, in which performances across each one are significantly correlated with all of the other three. This could stem from various different possibilities, for example, either that these skills all draw on the same underlying process(es) impacted in LTs or simply that this finding results from the inclusion of the LTs in the sample analysed in Publication I (Appendix A: Crawshaw et al., 2025b) especially if LTs could be differentially utilising underlying processes in an adaptive or compensatory manner. The decision that they should remain in this sample was made to represent natural variance in the population. The aim was to maintain comparability with other work investigating cross-sectional connections between gesture, language, ToM, and metaphor comprehension. Such work would not have had this detailed historical data to inform decision-making about any potential exclusions. Arbitrarily removing them and reanalysing this sample would significantly reduce its statistical power and generalisability as well as artificially restrict the data's ability to reflect a natural spectrum of ability (see Publication I, Appendix A: Crawshaw et al., 2025b). Future work should thus address the findings of Publications I and II (Crawshaw et al., 2025a, 2025b; Appendices A and B) within a much larger sample.

If we refer back to Figure 1 and the results of the stepwise regression and mediation analyses reported in Publication I (Appendix A: Crawshaw et al., 2025b), then index-finger pointing at 1;0 was directly predictive of two of the skills impacted in LTs (productive vocabulary) and indirectly predictive of the other two (ToM Strange Stories and receptive grammar) via the mediator composite language skills at age 3;0. Via the other mediation pathway of iconic gesture comprehension at 3;0, it was only indirectly predictive of receptive grammar. Considering these findings in relation to those regarding the LTs could signal various possibilities. Evidence from prior work demonstrates that gesture and spoken language undergo different developmental relations and trajectories in LTs (including those who go on to receive a diagnosis of DLD; e.g. Capone & McGregor, 2004; Guidetti & Nicoladis, 2008). This may be why we see a difference between the predictive or mediational capacities of index-finger pointing at 1;0, composite language skills at 3;0 and iconic gesture comprehension at 3;0 for the outcome skills at 9;0. Other work has shown that LTs initially produce fewer gestures than their TD peers but that this inverts as they become older and they ultimately use gesture more frequently to compensate for communicative deficits, albeit with lower accuracy and sustained weakness in their comprehension (c.f. Iverson et al., 2018; Lüke, Ritterfeld et al., 2017; Mainela-Arnold et al., 2014; Manwaring et al., 2019; Sansavini et al., 2021; Thal et al., 2013; Wray et al., 2016, 2017). This combination of both increased but less accurate use and comprehension of gestures in LTs may be why iconic gesture comprehension is a less effective mediator at 3;0 than composite language skills, especially if individual differences are exerting influence. Index-finger pointing at 1;0 may however be a clearer distinguisher of LT and TD children (c.f. Lüke, Grimminger et al., 2017; Lüke, Ritterfeld et al., 2017; however, for a meta-analysis of the relation between infant pointing and language development, see also Kirk et al., 2022).

Another noteworthy point is Thal et al.'s (2013) identification of a graded distinction in gesture use between subtypes of LTs, which they proposed was a reflection of different levels of representational abilities. From 10 months of age, TD children in their longitudinal sample used more gestures (both early communicative, non-symbolic as well as later representational ones) than 'late producers' (with a productive but not receptive deficit), who in turn used more than the 'late comprehenders' (with both productive and receptive deficits; Thal et al., 2013, p.179–180). The 'late producers' were more likely to 'recover', i.e. go on to perform within the TD range, while the 'late comprehenders' were more likely to remain delayed, i.e. receive a later DLD diagnosis (Thal et al., 2013). It may however be the case that the LTs presented in Publication II (Appendix B: Crawshaw et al., 2025a) do not clearly fit into the subtypes reported by Thal et al. (2013). Prior work has importantly highlighted that binary classifications are not applicable to LTs given their considerable individual differences in performance (Dollaghan, 2013). Unfortunately only eight LTs participated in the study reported in Publication II (Appendix B: Crawshaw et al., 2025a) and the results are thus not generalisable without extensive further research being carried out. However, it may still be of value to take a closer look at some of the children on an individual basis.

Figure 3

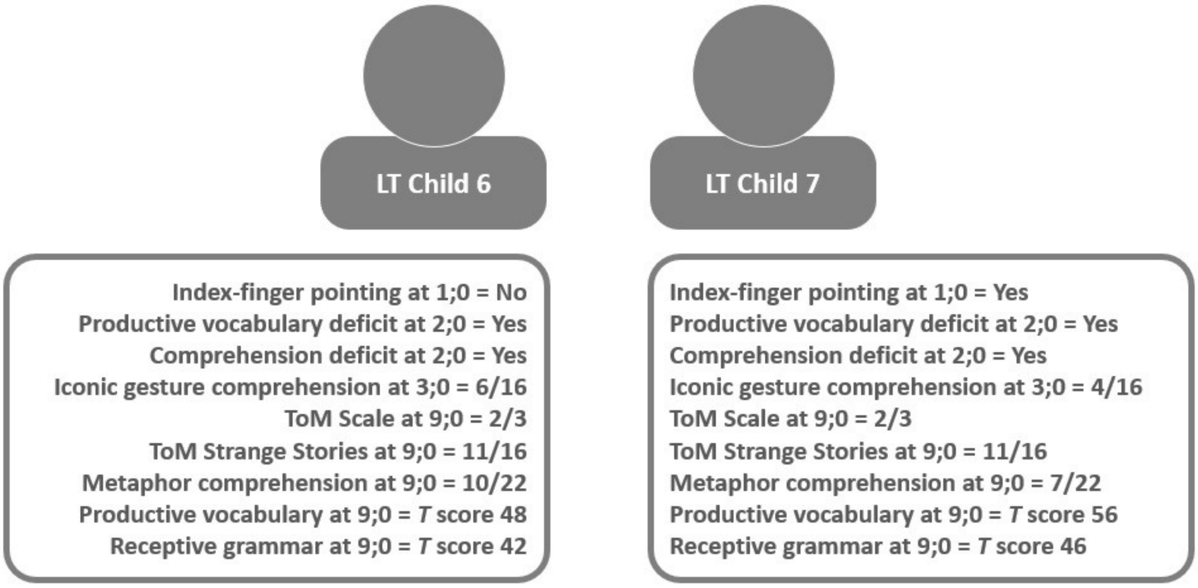
Visual representation of two LTs' performances across measures (LT Child 5 and LT Child 8; data partially reported in Appendix B: Crawshaw et al., 2025a).



Two cases particularly worth attention are that of LT Child 5 and LT Child 8 (Figure 3; data partially presented in Table 3 of Publication II, Appendix B: Crawshaw et al., 2025a). Child 5 (female) pointed with the index finger at age 1;0 while Child 8 (male) did not. Both had a productive vocabulary deficit but not a comprehension deficit at age 2;0 (i.e. were 'late producers' and not 'late comprehenders', c.f. Thal et al., 2013). At age 3;0, Child 8 only correctly understood 1 of the 16 iconic gestures tested, while Child 5 correctly understood 11; only 2 other children across the entire sample (including the TD group) had scored higher than her. At age 9;0, Child 5 had both the highest metaphor comprehension score (19 out of 22) and highest productive vocabulary score (T-Score of 68) across the whole sample (both TD and LT children). In contrast, Child 8 had the lowest metaphor comprehension score (1 out of 22) and third lowest productive vocabulary score (T-Score of 40) across the entire sample at age 9;0. They achieved the same receptive grammar T-Score (of 53) at 9;0 and their performances were also relatively comparable on the ToM Strange Stories task (The asterisk in Figure 3 refers to the fact that one Strange Story was accidentally omitted by the experimenter with Child 8; his score was thus not used for the analyses involving the ToM Strange Stories measure reported in Publications I and II: Crawshaw et al., 2025a, 2025b; Appendices A and B). Neither Child 5 nor Child 8 had a comprehension deficit at age 2;0 (both would potentially be classed under the subtype of 'late producer', c.f. Thal et al., 2013) so this does not provide an explanation for their differences in iconic gesture comprehension and metaphor comprehension, especially since their receptive grammar and ToM Strange Stories scores are fairly comparable. In search of a pattern, one could question whether their striking differences in index-finger pointing at 1;0 and iconic gesture comprehension at 3;0 might be linked to Child 5's significantly greater productive vocabulary and metaphor comprehension scores at 9;0.

Figure 4

Visual representation of two LTs' performances across measures (LT Child 6 and LT Child 7; data partially reported in Appendix B: Crawshaw et al., 2025a).



However, this hypothetical picture becomes less clear once you also consider the cases of LT Child 6 and LT Child 7 (Figure 4; data partially presented in Table 3 of Publication II, Appendix B: Crawshaw et al., 2025a). At age 2;0 they both had production and comprehension deficits, suggesting they might both fall under the subtype of 'late comprehenders' (c.f. Thal et al., 2013). The most noticeable difference between them is that Child 7 (male) was an index-finger pointer at 1;0 while Child 6 (female) was not; Child 5 (Figure 3) and Child 7 were actually the only index-finger pointing LTs in the sample. Despite this, Child 6 understood slightly more iconic gestures than Child 7 at age 3;0, and at age 9;0 she again understood slightly more metaphors than him. However Child 7 scored higher than Child 6 on both the productive vocabulary and receptive grammar measures at age 9;0. These performances (Child 6 and Child 7) do not fit with the hypothetical pattern emerging from the comparison between Child 5 and Child 8; we instead see a division where index-finger pointing capacity and productive vocabulary might be linked while iconic gesture comprehension and metaphor comprehension might have a separate connection. However, both the iconic gesture and metaphor comprehension tasks constituted items drawn from different classes (perceptual vs. pantomime iconic gestures, 11 metaphor categories; c.f. Publication I, Appendix A: Crawshaw et al., 2025b). They may thus have scored differently across the categories. Given that these abilities have also been linked to other skills within embodied cognition and relational thought paradigms, it is still possible that the children may have been utilising different skills in processing and achieving understanding of specific different gestures or metaphors. It might be instructive if new tasks could be designed to identify to how LTs perform comparably to TD children on tasks evaluating their understanding of shared intentionality and relational thinking skills. Certain spatial cognition tasks might also be able to provide insights into a participant's ability to mentally visualise (represent an embodiment of) themselves and embed this within certain contexts to solve problems. Assessing these, however, would require further in-depth research and analysis beyond the scope of this dissertation. Nevertheless, these four cases (LT Children 5, 6, 7, and 8) highlight the significant amount of complexity involved in even beginning to disentangle the relations between skills once you start to account for the great extent of variance in their individual differences.

Instead of categorising LTs under specific subtypes, it may be more helpful to plot them along a spectrum. Across Publication II (Appendix B: Crawshaw et al., 2025a), the potential existence of a TD-LT-DLD spectrum was repeatedly proposed. This hypothetical spectrum would align with and extend prior ideas regarding a “language endowment spectrum” (c.f. Ellis Weismer, 2007, p. 84) as well as the idea that LTs may fall somewhere between TD and DLD children along dimensional spectra of language(-related) abilities (Rescorla, 2009, 2011; Thal et al., 2013). Given the longitudinal associations reported between gesture, language, ToM, and metaphor comprehension (all language-related skills) in Publication I (Appendix A: Crawshaw et al., 2025b), it may be helpful to further conceptualise such a TD-LT-DLD spectrum as existing within a developmental continuum of language-related skills. Some have connected LTs to difficulties with early representational skills (e.g. Thal et al., 2013). The domains of ToM and metaphor comprehension draw on metarepresentational, metacognitive, and metalinguistic skills and are also strongly associated with general language skills (impacted in LTs). If representational (and potentially relational) skills are also affected in LTs, then logically these domains hold great potential for future, more fine-grained exploration in LTs. Since the index-finger pointing measure used in Publication I (Appendix A: Crawshaw et al., 2025b) received a binary yes or no score, it would be important for future research to work with a more nuanced measure of pointing in order to integrate it into such a proposed continuum.

In summary, the findings of Publication I (Appendix A: Crawshaw et al., 2025b) indicate the existence of a complex, interrelated network of language-related skills which may draw upon common, underlying cognitive processes. The findings of Publication II (Appendix B: Crawshaw et al., 2025a) suggest that some but not all of these skills are impacted in LTs, which may imply that specific common underlying processes (or areas of processing) within this network may be impacted in terms of delay or deficit. This should be addressed further in future work, especially with regard to metacognitive and pragmatic skills such as ToM and metaphor comprehension, which prior work had ostensibly not addressed in LTs without a diagnosis of DLD.

6.3 Limitations

Naturally, there are some limitations to the work presented in this dissertation which must also be addressed. Certain limitations are already reported and discussed in detail within Publications I and II (Crawshaw et al., 2025a, 2025b; Appendices A and B) which include (1) the small total sample size of participants (N = 35), (2) the inclusion of the LTs in the wider analyses, (3) the uneven group sizes of participants (TD n = 27; LT n = 8) in the comparative analyses which affect the generalisability of the findings, and (4) the hybrid digital and in-person experimental procedure which had to be developed in response to the concurrent COVID-19 pandemic.

There are however two further general limitations relating to both Publications I and II (Crawshaw et al., 2025a, 2025b; Appendices A and B) which should also be acknowledged and evaluated. Firstly, certain skill domains were not addressed. Like receptive vocabulary (discussed in section 6.1), iconic and metaphoric gesture production and comprehension were not examined at age 9;0 in relation to the other skills considered. These would have been interesting to consider and may have facilitated broader exploration of the potential interrelations between the skills. By focusing solely on their comprehension, the work supporting this dissertation only provides partial insight into the connections between iconic gesture and metaphor. Since iconic gesture and metaphor comprehension appear to be related differentially to specific language skills, this might also be true for their production.

In a similar vein, it might also have been of value to integrate measures of executive function (EF) and working memory into the analyses, since ToM and EF have been both associated and disassociated in past work (c.f. Apperly, 2021; Devine, 2021; Tomasello, 2019). Processing ToM and complex language simultaneously involves EF and working memory costs (Apperly, 2021). Children with a better command of certain EF skills might thus have performed better across certain tasks since different cognitive processes are utilised to infer, store, and apply mentalising information, which may not all be necessary within every situation (Apperly, 2021; however, see also Devine, 2021). For example, processes of cognitive reorganisation and representational redescription (to facilitate increasingly abstracted understanding and coordinate perspectives; Tomasello, 2019, p.36) would be important for each of gesture, language, ToM, and metaphor.

Secondly, there are some limitations to the ToM tasks which were selected. One inherent problem concerning many ToM tasks (including those used within the publications supporting this dissertation) is that they present artificial singletons or dyads to a participant when in real life they have to take account of and integrate many different people's mental states simultaneously (Lagattuta & Kramer, 2021). Performance on these tasks may therefore not provide an accurate reflection of a participant's true ToM abilities.

Another issue is that all individuals, regardless of age, vary in the extent to which they engage in mindreading (c.f. Devine, 2021); this could also be influenced by contextual factors such as concurrent environmental challenges or an individual's emotional regulatory status. Children are still developing their ToM capacities and accounting for their individual differences could be even more complex: Differences in performance may also result from an incomplete understanding of mental-state concepts, inadequate EF ability to enable them to change focus onto another person's perspective, or insufficient language skills to verbalise their understanding of other's minds (Devine, 2021, p.55). Another related limitation to this dissertation's supporting work is that ToM had not been tested at an earlier point in the wider longitudinal study so there was no measure of the children's heterotypic stability (c.f. Devine, 2021, p.64). Having an earlier measure of ToM would have meant that differences between individuals could have been evaluated on the basis of whether they performed consistently across age points compared to the wider group. Potential relations between gesture and ToM could also then have been evaluated at an earlier age point. This would have subsequently facilitated observation of how this relation might change over time or identification of differences in the underlying constructs measured. However, Devine

(2021) conducted a meta-analysis indicating that individual differences in ToM in early and middle childhood appear moderately stable even across different measures. For this reason, the results of an earlier measure of ToM would likely have been relatively consistent with those of the ToM measures conducted at age 9;0.

The design of the ToM Scale task means that it may also not have been the most effective test for measuring ToM in children aged 9;0. The ToM Scale (e.g. Wellman & Liu, 2004) is designed to identify early vs. late mastery of certain ToM concepts and is not particularly sensitive to performance differences within a specific age group (Devine, 2021). Children typically pass specific items in a predictable order and standard false belief tasks usually do not adequately indicate individual ToM differences after age 6 due to the presence of ceiling effects (Devine, 2021, p.58, p.59). Accordingly, the original decision to use this test at age 9;0 had actually been made to establish a baseline measure of abilities. Surprisingly, even at age 9;0 no ceiling effects were found across any of the ToM Scale subtests used in Publications I and II (Crawshaw et al., 2025a, 2025b; Appendices A and B; N = 34; hidden emotions: pass n = 27, fail n = 7; sarcasm: pass n = 14, fail n = 20; explicit false belief: pass n = 30, fail n = 4). Four children did not pass the explicit false belief subtest (only one of them was a LT) and three of those four actually still passed the harder sarcasm subtest (the LT did not). This indicates that ToM concept acquisition is not necessarily a linear process.

Despite work that has identified coherence and consistency across ToM tasks (c.f. Devine, 2021), the children's performances on the ToM Scale and ToM Strange Stories tasks were not correlated (see Publication I, Appendix A: Crawshaw et al., 2025b). Work with children in middle childhood has also found low levels of internal consistency between some sets of advanced ToM tasks (c.f. Devine, 2021). However, latent variable modelling methods have shown that performance on the Strange Stories task loads onto one latent factor also shared by other tasks requiring children to explain behaviour by referring to mental states (Devine, 2021). Devine (2021) proposed that a lack of coherence between advanced ToM measures may instead reflect (a) variance in their discriminating power to identify between-participant differences across the age range, (b) statistical inconsistencies resulting from sample size and composition, or even (c) task impurity problems arising from unique task demands which mask relations. Any of these could provide further explanations for the absence of a correlation between the ToM tasks selected for Publications I and II (Crawshaw et al., 2025a, 2025b; Appendices A and B). These issues can only be detangled in future work addressing larger sample sizes as well as broader selections of ToM measures. In general, however, the use of the ToM Strange Stories task is a strength of the work presented here since it has been shown to depict substantial variance across participants in middle childhood and beyond (c.f. Devine, 2021). In comparison to some other ToM tasks, the Strange Stories task has also been recognised for having higher face validity and stronger measurement characteristics (Apperly, 2021). One limitation is that it nevertheless appears to provide less insight into underlying cognitive processes (Apperly, 2021), which reduces opportunities to evaluate the potential subnetworks presented in Figures 1 and 2.

A final factor is the composition of the ToM Scale task in addition to the timing of its use. The ToM Scale (originally by Wellman & Liu, 2004, Peterson et al., 2012; translated by Henning et al., 2013) is labelled for use with children aged 3–11 and comprises four further subtests (not-own desire, not-own belief, knowledge access, and false belief unexpected contents) which were not conducted in the work presented here. In the work supporting this dissertation, the decision was made to select three of the subtests for age-appropriateness to fit within the time constraints of the experimental sessions, given that so many other measures were being collected concurrently. In hindsight, since ceiling effects were not found in any of the 3 selected subtests, it might have been better to have conducted all 7 subtests. This could have improved the ability of the total score (0–7 instead of 0–3) to reflect variance among the participating children and better facilitated analysis of its links to other related skills. Another minor issue is that the recommended order of subtests was not followed since only specific subtests had been selected for use, which could have influenced the results. The same order was however followed across all participants.

The sarcasm subtest (Peterson et al., 2012) has been recognised as the most challenging of the ToM Scale and the best suited for capturing variation in task performance in middle childhood (c.f. Devine, 2021). This was also reflected in the results of the work presented here (pass n = 14, fail n = 20; only 2 of the 8 LTs passed this subtest, these were LT Child 5, see Figure 3, and LT Child 4, see Publication II, Appendix B: Crawshaw et al., 2025a). Prior work indicates that an understanding of sarcasm or irony develops relatively late, emerging only from around the age of 6 years, prior to which it is typically interpreted literally (c.f. Falkum, 2019, p.17). Falkum (2019, p.6) proposed that in contrast to metaphors, comprehending irony requires recognising that the speaker has distanced themselves from their utterance's literal content, necessitating second-order metarepresentational skills (Falkum, 2019, p.17). Falkum and Köder (2020, p.6) also identified that children's emerging understanding of irony at age 6 coincides with a period of significant progression in higher order ToM development. However three children in the sample presented here passed the sarcasm subtest but failed the explicit FB subtest. This could be an idiosyncrasy of this dataset (i.e. at that moment, on that day, this task may not have accurately captured those individuals' true ToM capacities), meaning caution may be needed in the interpretation of the findings presented here. However the possibility of an alternative route to irony comprehension could be considered in future work, especially given that the explicit FB task was supported visually and used language with very low processing demands. Those three children did not reach ceiling effects across all the items of the ToM Strange Stories task (which involve higher order metarepresentation), implying that their acquisition of certain ToM concepts was not yet well established. As the majority of children in the dataset failed this task, this may raise questions about timing of its use. It would be helpful to have been able to conduct additional measures of sarcasm comprehension both earlier and in future work with the children to gain an understanding of when it emerged in individuals relative to their other competencies. A longitudinal approach to analysing and considering when children begin to pass sarcasm and irony comprehension tasks in relation to their concurrent levels of other linguistic, metalinguistic, pragmatic, and metacognitive skills would be valuable and interesting to see in future work.

7 Conclusions and Future Directions

The work presented here in support of this dissertation highlights the interwoven nature of and the complexity in relations between the domains of gesture, spoken language, ToM, and metaphor comprehension. Cognitive functions rely on complex underlying networks and abilities are not exercised in an isolated manner. Gesture is a developmental precursor and important predictor of spoken language abilities, constituting a key component of interpersonal communication across the life span. Gesturing with our hands allows us to refer to and represent concrete and abstract concepts as well as internal and external events and entities. However, to produce and comprehend these gestures, we must utilise, coordinate, and integrate myriad motor, visual, perceptual, conceptual, cognitive, and representational skills. These in turn rely on wider processes such as decoding, inference-making, and relational thinking as well as on having achieved an understanding of conventionalisation and recursive intentionality (both in its joint and collective forms; see Tomasello, 2019). Across development, these same skills are also critically important for the acquisition and comprehension of spoken language, ToM, and metaphor.

Along with their gestural and language skills, children also vary in their metaphor comprehension and ToM abilities. This has a real-world impact on their social experiences, relationships, and outcomes, affecting their internalised perspectives of themselves in relation to others (c.f. Del Sette et al., 2021; Devine, 2021; Tomasello, 2019). Regardless of whether a developmental continuum of language-related skills exists, language development appears to be supported by gestural competencies, and language abilities in turn seem to facilitate ToM and metaphor comprehension. Atypically developing populations (such as LTs) struggling with any of language production, comprehension, and processing are thus at risk of knock-on delays within these other language-associated domains. Future work should integrate measures of gesture into any consideration of potential relations between spoken language, ToM, and metaphor comprehension in both typically and atypically developing populations. The links found between gesture and both ToM and metaphor comprehension in the work presented in this dissertation also signify interesting opportunities for generating new teaching and intervention methods. Researchers aiming to develop potential interventions for at-risk groups might consider creating gesture-assisted tasks for training ToM or metaphor comprehension as well as comparing their effectiveness with spoken language versions. If gesture, ToM, and metaphor all share roots in relational thinking as well as embodied and embedded cognition, then children's developing familiarity with and processing of them could be facilitated by highlighting relations between relevant concepts via embodied gestures and embedding these within appropriate training contexts.

8 References

Andrés-Roqueta, C., Adrian, J. E., Clemente, R. A., & Katsos, N. (2013). Which are the best predictors of theory of mind delay in children with specific language impairment?. *International Journal of Language & Communication Disorders*, *48*(6), 726–737. https://doi.org/10.1111/1460-6984.12045

Apperly, I. A. (2021). Cognitive basis of mindreading in middle childhood and adolescence. In R. T. Devine & S. Lecce (Eds.), *Theory of Mind in Middle Childhood and Adolescence* (pp. 37–54). Routledge. https://doi.org/10.4324/9780429326899

Armstrong, R. A. (2014). When to use the Bonferroni correction. *Ophthalmic and Physiological Optics, 34*(5), 502–508. https://doi.org/10.1111/opo.12131

Astington, J. W. & Baird, J. A. (2005). Introduction: Why Language Matters. In J. W. Astington & J. A. Baird (Eds.), *Why Language Matters for Theory of Mind* (pp.3–25). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195159912.003.0001

Atkinson, L., Slade, L., Powell, D., & Levy, J. P. (2017). Theory of mind in emerging reading comprehension: A longitudinal study of early indirect and direct effects. *Journal of Experimental Child Psychology*, *164*, 225–238. https://doi.org/10.1016/j.jecp.2017.04.007

Babarczy, A., Balázs, A., & Krizsai, F. (2019). Preschoolers' Metaphor Comprehension. Methodological Issues in Experimental Pragmatics. *Acta Universitatis Sapientiae, Philologica*, *11*(2), 133–150. https://doi.org/10.2478/ausp-2019-0017

Baggio, G. (2025). Gesture, meaning, and intentionality: from radical to pragmatist enactive theory of language. *Phenomenology and the Cognitive Sciences, 24*(1), 33–62. https://doi.org/10.1007/s11097-023-09936-9

Barsalou, L. W., & Wiemer-Hastings, K. (2005). Situating abstract concepts. In D. Pecher & R. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thought* (pp. 129–163). Cambridge University Press.

Bates, E., Camaioni, L., & Volterra, V. (1975). The Acquisition of Performatives Prior to

Speech. *Merrill-Palmer Quarterly of Behavior and Development, 21*(3), 205–226. http://www.jstor.org/stable/23084619

Bates, E., Dale, P., & Thal, D. (1995). Individual differences and their implications for theories of language development. In P. Fletcher & B. MacWhinney (Eds.), *Handbook of Child Language* (pp. 96–151). Basil Blackwell. https://doi.org/10.1111/b.9780631203124.1996.00005.x

Bates, E., Thal, D., Whitesell, K., Fenson, L., & Oakes, L. (1989). Integrating language and gesture in infancy. *Developmental Psychology, 25*(6), 1004–1019. https://doi.org/10.1037/0012-1649.25.6.1004

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological), 57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics, 29*(4) 1165–1188. http://www.jstor.org/stable/2674075

Bigelow, F. J., Clark, G. M., Lum, J. A., & Enticott, P. G. (2021). The mediating effect of language on the development of cognitive and affective theory of mind. *Journal of Experimental Child Psychology, 209*, 105158. https://doi.org/10.1016/j.jecp.2021.105158

Bishop, D. V. M. (2017). Why is it so hard to reach agreement on terminology? The case of developmental language disorder (DLD). *International Journal of Language & Communication Disorders, 52*(6), 671–680. https://doi.org/10.1111/1460-6984.12335

Bishop, D. V. M., & Edmundson, A. (1987). Language-impaired 4-year-olds: Distinguishing transient from persistent impairment. *Journal of Speech and Hearing Disorders, 52*(2), 156–173. https://doi.org/10.1044/jshd.5202.156

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & Catalise-2 Consortium. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry, 58*(10), 1068–1080. https://doi.org/10.1111/jcpp.12721

Bland, J. M., & Altman, D. G. (1995). Multiple significance tests: the Bonferroni method. *BMJ, 310*(6973), 170. https://doi.org/10.1136/bmj.310.6973.170

Boerma, I. E., Mol, S. E., & Jolles, J. (2017). The role of home literacy environment, mentalizing, expressive verbal ability, and print exposure in third and fourth graders' reading comprehension. *Scientific Studies of Reading, 21*, 179–193. https://doi.org/10.1080/10888438.2016.1277727

Borleffs, E., Maassen, B. A., Lyytinen, H., & Zwarts, F. (2019). Cracking the code: The impact of orthographic transparency and morphological-syllabic complexity on reading and developmental dyslexia. *Frontiers in Psychology, 9*, 2534. https://doi.org/10.3389/fpsyg.2018.02534

Botting, N., Riches, N., Gaynor, M., & Morgan, G. (2010). Gesture production and comprehension in children with specific language impairment. *British Journal of Developmental Psychology*, *28*(1), 51–69. https://doi.org/10.1348/026151009X482642

Brandone, A. C., & Stout, W. (2023). The Origins of Theory of Mind in Infant Social Cognition: Investigating Longitudinal Pathways from Intention Understanding and Joint Attention to Preschool Theory of Mind. *Journal of Cognition and Development, 24*(3), 375–396. https://doi.org/10.1080/15248372.2022.2146117

Braun, R., Blok, V., Loeber, A., & Wunderle, U. (2020). COVID-19 and the onlineification of research: kick-starting a dialogue on Responsible online Research and Innovation (RoRI). *Journal of Responsible Innovation*, *7*(3), 680–688. https://doi.org/10.1080/23299460.2020.1789387

Bühler, D., Perovic, A., & Pouscoulous, N. (2018). Comprehension of novel metaphor in young children with Developmental Language Disorder. *Autism & Developmental Language Impairments*, *3*, 1–11. https://doi.org/10.1177/2396941518817229

Camaioni, L. (1992). Mind knowledge in infancy: The emergence of intentional communication. *Early Development and Parenting*, *1*(1), 15–22. https://doi.org/10.1002/edp.2430010106

Camaioni, L., Perucchini, P., Bellagamba, F., & Colonnesi, C. (2004). The role of declarative pointing in developing a theory of mind. *Infancy*, *5*(3), 291–308. https://doi.org/10.1207/s15327078in0503_3

Capone, N.C., McGregor, K.K., 2004. Gesture development: a review for clinical and research practices. *Journal of Speech, Language, and Hearing Research, 47*(173), 173–186. https://doi.org/10.1044/10924388(2004/015)

Carriedo, N., Corral, A., Montoro, P. R., Herrero, L., Ballestrino, P., & Sebastián, I. (2016). The development of metaphor comprehension and its relationship with relational verbal reasoning and executive function. *PloS One*, *11*(3), e0150289. https://doi.org/10.1371/journal.pone.0150289

Carlson, S. M., Wong, A., Lemke, M., & Cosser, C. (2005). Gesture as a window on children's beginning understanding of false belief. *Child Development*, *76*(1), 73–86. https://doi.org/10.1111/j.1467-8624.2005.00830.x

Chahboun, S., Vulchanov, V., Saldaña, D., Eshuis, H., & Vulchanova, M. (2016). Can you play with fire and not hurt yourself? A comparative study in figurative language comprehension between individuals with and without autism spectrum disorder. *PloS One*, *11*(12), e0168571. https://doi.org/10.1371/journal.pone.0168571

Cienki, A., & Müller, C. (2008). Metaphor, gesture, and thought. In R. W. J. Gibbs (Ed.), *The Cambridge Handbook of Metaphor and Thought* (pp. 483–501). Cambridge University Press.

Cochet, H., Jover, M., Rizzo, C., & Vauclair, J. (2017). Relationships between declarative pointing and theory of mind abilities in 3- to 4-year-olds, *European Journal of Developmental Psychology, 14*(3), 324–336.

https://doi.org/10.1080/17405629.2016.1205975

Colonnesi, C., Rieffe, C., Koops, W., & Perucchini, P. (2008). Precursors of a theory of mind: A longitudinal study. *British Journal of Developmental Psychology*, *26*(4), 561–577. https://doi.org/10.1348/026151008X285660

Colonnesi, C., Stams, G. J. J., Koster, I., & Noom, M. J. (2010). The relation between pointing and language development: A meta-analysis. *Developmental Review*, *30*(4), 352–366. https://doi.org/10.1016/j.dr.2010.10.001

Cooperrider, K., & Goldin-Meadow, S. (2017). When gesture becomes analogy. *Topics in Cognitive Science, 9*(3), 719–737. https://doi.org/10.1111/tops.12276

Crawshaw, C. E., Lüke, C., & Ritterfeld, U. (2025a). Late Talkers' Language, Metaphor, Theory of Mind, and Reading Skills at 9 Years. *Journal of Speech, Language, and Hearing Research*, *68*(3), 1038–1055. https://doi.org/10.1044/2024_JSLHR-23-00703

Crawshaw, C. E., Lüke, C., & Ritterfeld, U. (2025b). Does early gesture usage contribute alongside oral language to later theory of mind performance and metaphor comprehension? Indications from a longitudinal study with children aged 1–9 years. *Language Acquisition*, 1–31. https://doi.org/10.1080/10489223.2025.2455174

Crawshaw, C. E., Kern, F., Mertens, U. and Rohlfing, K. J. (2020). Children's Narrative Elaboration After Reading a Storybook Versus Viewing a Video. *Frontiers in Psychology, 11:569891*, 1–16. https://doi.org/10.3389/fpsyg.2020.569891

Dargue, N., Sweller, N., & Jones, M. P. (2019). When our hands help us understand: A meta-analysis into the effects of gesture on comprehension. *Psychological Bulletin, 145*(8), 765–784. https://doi.org/10.1037/bul0000202

Davidson, R., & MacKinnon, J. G. (1993). Estimation and Inference in Econometrics. Oxford University Press.

Deckert, M., Schmoeger, M., Schaunig-Busch, I., & Willinger, U. (2019). Metaphor processing in middle childhood and at the transition to early adolescence: The role of chronological age, mental age, and verbal intelligence. *Journal of Child Language*, *46*(2), 334–367. https://doi.org/10.1017/S0305000918000491

Del Sette, P., Bambini, V., Bischetti, L., & Lecce, S. (2020). Longitudinal associations between theory of mind and metaphor understanding during middle childhood. *Cognitive Development*, *56*, 100958. https://doi.org/10.1016/j.cogdev.2020.100958

Del Sette, P., Ronchi, L., Bambini, V., & Lecce, S. (2021). Longitudinal associations between metaphor understanding and peer relationships in middle childhood. *Infant and Child Development*, e2232. https://doi.org/10.1002/icd.2232

Demir, Ö. E., Levine, S. C., & Goldin-Meadow, S. (2015). A tale of two hands: Children's early gesture use in narrative production predicts later narrative structure in speech. *Journal of Child Language, 42*(3), 662–681. https://doi.org/10.1017/S0305000914000415

De Mulder, H. N. M., Wijnen, F., & Coopmans, P. H. A. (2019). Interrelationships between Theory of Mind and language development: A longitudinal study of Dutch-speaking kindergartners. *Cognitive Development*, *51*, 67–82. https://doi.org/10.1016/j.cogdev.2019.03.006

De Rosnay, M., & Hughes, C. (2006). Conversation and theory of mind: Do children talk their way to socio-cognitive understanding?. *British Journal of Developmental Psychology, 24*(1), 7–37. https://doi.org/10.1348/026151005X82901

Deschrijver, E., & Palmer, C. (2020). Reframing social cognition: Relational versus representational mentalizing. *Psychological Bulletin, 146*(11), 941–969. https://doi.org/10.1037/bul0000302

De Villiers, J. G. (2007). The interface of language and theory of mind. *Lingua, 117*(11), 1858–1878. https://doi.org/10.1016/j.lingua.2006.11.006

De Villiers, J. G., & Pyers, J. E. (2002). Complements to cognition: A longitudinal study of the relationship between complex syntax and false-belief-understanding. *Cognitive*

Development, 17(1), 1037–1060. https://doi.org/10.1016/S0885-2014(02)00073-4

De Villiers, P. A. (2005). The Role of Language in Theory-of-Mind Development: What Deaf Children Tell Us. In J. W. Astington & J. A. Baird (Eds.), *Why Language Matters for Theory of Mind* (pp. 266–297). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195159912.003.0013

Devine, R. T. (2021). Individual differences in theory of mind in middle childhood and adolescence. In R. T. Devine & S. Lecce (Eds.), *Theory of Mind in Middle Childhood and Adolescence* (pp. 55–76). Routledge. https://doi.org/10.4324/9780429326899

Devine, R. T., & Hughes, C. (2016). Measuring theory of mind across middle childhood: Reliability and validity of the silent films and strange stories tasks. *Journal of Experimental Child Psychology*, *149*, 23–40. https://doi.org/10.1016/j.jecp.2015.07.011

Devine, R. T., & Lecce, S. (2021). Why study theory of mind in middle childhood and adolescence? In R. T. Devine & S. Lecce (Eds.), *Theory of Mind in Middle Childhood and Adolescence* (pp. 1–11). Routledge. https://doi.org/10.4324/9780429326899

Devine, R. T., White, N., Ensor, R., & Hughes, C. (2016). Theory of mind in middle childhood: Longitudinal associations with executive function and social competence. *Developmental Psychology*, *52*(5), 758–771. https://doi.org/10.1037/dev0000105

Diamanti, V., Goulandris, N., Campbell, R., & Protopapas, A. (2018). Dyslexia profiles across orthographies differing in transparency: An evaluation of theoretical predictions contrasting English and Greek. *Scientific Studies of Reading, 22*(1), 55–69. https://doi.org/10.1080/10888438.2017.1338291

Di Paola, S., Domaneschi, F., & Pouscoulous, N. (2020). Metaphorical developing minds: The role of multiple factors in the development of metaphor comprehension. Journal of Pragmatics, 156, 235–251. https://doi.org/10.1016/j.pragma.2019.08.008

Doğan, I., Özer, D., Aktan-Erciyes, A., Furman, R., Demir-Lira, Ö. E., Özçalışkan, Ş., &

Göksun, T. (2024). The link between early iconic gesture comprehension and receptive language. *Infant and Child Development*, e2552. https://doi.org/10.1002/icd.2552

Doherty, M. (2008). *Theory of mind: How children understand others' thoughts and feelings*. Psychology Press. https://doi.org/10.4324/9780203929902

Doherty, M. (2000). Children's understanding of homonymy: Metalinguistic awareness and false belief. *Journal of Child Language*, *27*(2), 367–392. https://doi.org/10.1017/S0305000900004153

Doherty, M., & Perner, J. (2020). Mental files: Developmental integration of dual naming and theory of mind. *Developmental Review*, *56*, 100909. https://doi.org/10.1016/j.dr.2020.100909

Doherty, M., & Perner, J. (1998). Metalinguistic awareness and theory of mind: Just two words for the same thing? *Cognitive Development*, *13*(3), 279–305. https://doi.org/10.1016/S0885-2014(98)90012-0

Dollaghan, C. (2013). Late talker as a clinical category: A critical evaluation. In L. Rescorla & P. S. Dale (Eds.), *Late talkers: Language development, interventions, and outcomes* (pp. 91–112). Brookes.

Dove, G., Barca, L., Tummolini, L., & Borghi, A. M. (2022). Words have a weight: Language as a source of inner grounding and flexibility in abstract concepts. *Psychological Research, 86*(8), 2451–2467. https://doi.org/10.1007/s00426-020-01438-6

Durrleman, S. (2020). Mentalizing: What's language got to do with it? *Language Acquisition, 27*, 255–275. https://doi.org/10.1080/10489223.2020.1769624

Durrleman, S., Burnel, M., & Reboul, A. (2017). Theory of mind in SLI revisited: links with syntax, comparisons with ASD. *International Journal of Language & Communication Disorders*, *52*(6), 816–830. https://doi.org/10.1111/1460-6984.12317

Durrleman, S., & Delage, H. (2020). Training Complements for Belief Reasoning in

Developmental Language Disorder. *Journal of Speech, Language, and Hearing Research*, *63*(6), 1861–1877. https://doi.org/10.1044/2020_JSLHR-19-00075

Ebert, S. (2020). Theory of mind, language, and reading: Developmental relations from early childhood to early adolescence. *Journal of Experimental Child Psychology*, *191*, 104739. https://doi.org/10.1016/j.jecp.2019.104739

Ensor, R., Devine, R. T., Marks, A., & Hughes, C. (2014). Mothers' cognitive references to 2-year-olds predict theory of mind at ages 6 and 10. *Child Development*, *85*(3), 1222–1235. https://doi.org/10.1111/cdev.12186

Ellis Weismer, S. (2007). Typical talkers, late talkers, and children with specific language impairment: A language endowment spectrum? In R. Paul (Ed.), *The influence of developmental perspectives on research and practice in communication disorders: A festschrift for Robin S. Chapman* (pp. 83–102). Erlbaum.

Escudero, P., Pino Escobar, G., Casey, C. G., & Sommer, K. (2021). Four-Year-Old's Online Versus Face-to-Face Word Learning via eBooks. *Frontiers in Psychology*, *12*, 450. https://doi.org/10.3389/fpsyg.2021.610975

Esser, G., Wyschkon, A., Ballaschk, K., & Hänsch, S. (2010). *Potsdam-Illinois Test für Psycholinguistische Fähigkeiten – P-ITPA* [Potsdam-Illinois Test for Psycholinguistic Abilities – P-ITPA]. Hogrefe, Testzentrale.

Falkum, I. L. (2019). Pragmatic development. Learning to use language to communicate. In J. S. Horst & J. von Koss Torkildsen (Eds.), *International Handbook of Language Acquisition* (pp. 234–260). Routledge.

Farrant, B. M. (2015). Specific language impairment and perspective taking: Delayed development of theory of mind, visual and emotional perspective taking. *Journal of Childhood & Developmental Disorders*, *1*(1), 1–8. https://doi.org/10.4172/2472-1786.100008

Farrant, B. M., Fletcher, J., & Maybery, M. T. (2006). Specific language impairment, theory of mind, and visual perspective taking: Evidence for simulation theory and the developmental role of language. *Child Development*, *77*(6), 1842–1853. https://doi.org/10.1111/j.1467-8624.2006.00977.x

Farrar, M. J., Johnson, B., Tompkins, V., Easters, M., Zilisi-Medus, A., & Benigno, J. P. (2009). Language and theory of mind in preschool children with specific language impairment. *Journal of Communication Disorders*, *42*(6), 428–441. https://doi.org/10.1016/j.jcomdis.2009.07.001

Fenici, M. (2012). Embodied social cognition and embedded theory of mind. *Biolinguistics, 6*(3–4), 276–307. https://doi.org/10.5964/bioling.8921

Fox, A. V. (2013). *TROG-D. Test zur Überprüfung des Grammatikverständnisses* [Test for Assessing Grammatical Understanding] 6th Edition. Schulz-Kirchner.

Garfield, J. L., Peterson, C. C., & Perry, T. (2001). Social cognition, language acquisition and the development of the theory of mind. *Mind & Language, 16*(5), 494–541. https://doi.org/10.1111/1468-0017.00180

Gillott, A., Furniss, F., & Walter, A. (2004). Theory of mind ability in children with specific language impairment. *Child Language Teaching and Therapy*, *20*(1), 1–11. https://doi.org/10.1191/0265659004ct260oa

Goldin-Meadow, S. (2020). Using gesture to identify and address early concerns about language and pragmatics. *Pediatrics*, *146*(Supplement 3), S278–S283. https://doi.org/10.1542/peds.2020-0242G

Goldin-Meadow, S. (2000). Beyond words: The importance of gesture to researchers and learners. *Child Development, 71*(1), 231–239. https://doi.org/10.1111/1467-8624.00138

Goldin-Meadow, S. (1998). The development of gesture and speech as an integrated system. *New Directions for Child Development*, *79*, 29–42. https://doi.org/10.1002/cd.23219987903

Gordon Pershey, M. (2000). Children's Elicited Use of Pragmatic Language Functions: How Six-and Seven-year-old Children Adapt to the Interactional Environments of Story Scenarios. *Language Awareness, 9*(4), 218–235. https://doi.org/10.1080/09658410008667147

Gray, L. M., Wong-Wylie, G., Rempel, G. R., & Cook, K. (2020). Expanding qualitative research interviewing strategies: Zoom video communications. *The Qualitative Report*, *25*(5), 1292–1301. https://doi.org/10.46743/2160-3715/2020.4212

Graziano, M., & Gullberg, M. (2018). When speech stops, gesture stops: Evidence from developmental and crosslinguistic comparisons. *Frontiers in Psychology*, *9*, 879. https://doi.org/10.3389/fpsyg.2018.00879

Grimm, H. (2000). *SETK-2. Sprachentwicklungstest für zweijährige Kinder* [Language development test for two-year-old children]. Hogrefe.

Grimm, H. (2001). *Sprachentwicklungstest für drei- bis fünfjährige Kinder (SETK 3–5): Diagnose von Sprachverarbeitungsfähigkeiten und auditiven Gedächtnisleistungen* [Language development test for three- to five-year-old children (SETK 3–5): diagnosis of language processing skills and auditory memory skills]. Hogrefe.

Grimminger, A., Rohlfing, K. J., Lüke, C., Liszkowski, U., & Ritterfeld, U. (2019). Decontextualized talk in caregivers' input to 12-month-old children during structured interaction. *Journal of Child Language, 47*(2), 418–434. https://doi.org/10.1017/S0305000919000710

Guidetti, M., & Nicoladis, E. (2008). Introduction to special issue: Gestures and communicative development. *First Language, 28*(2), 107–115. https://doi.org/10.1177/0142723708088914

Hammer, C. S., Morgan, P., Farkas, G., Hillemeier, M., Bitetti, D., & Maczuga, S. (2017). Late talkers: A population-based study of risk factors and school readiness consequences. *Journal of Speech, Language, and Hearing Research, 60*(3), 607–626.

https://doi.org/10.1044%2F2016_JSLHR-L-15-0417

Happé, F. G. E. (1995a). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development, 66*(3), 843–855. https://doi.org/10.1111/j.1467-8624.1995.tb00909.x

Happé, F. G. E. (1995b) Understanding Minds and Metaphors: Insights from the Study of Figurative Language in Autism. *Metaphor and Symbol, 10*(4), 275–295. https://doi.org/10.1207/s15327868ms1004_3

Happé, F. G. E. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders*, *24*(2), 129–154. https://doi.org/10.1007/BF02172093

Hayes, A. F. (2018). *Introduction to Mediation, Moderation, and Conditional Process Analysis: Methodology in the Social Sciences*, 2nd Edition. Guilford Press.

Hemmerich, W. (2016). StatistikGuru: Rechner zur Adjustierung des α-Niveaus [*StatisticGuru: Calculator for adjusting the α level*]. https://statistikguru.de/rechner/adjustierung-des-alphaniveaus.html

Henning, A., Hofer, T., & Aschersleben, G. (2013). *Erweiterte "Theory of Mind-Skala" für 3- bis 11-jährige Kinder* [Extended Theory-of-Mind Scale for 3- to 11-year-old children], 2nd Edition. Universität des Saarlandes Arbeitseinheit Entwicklungspsychologie.

Hjetland, H. N., Lervåg, A., Lyster, S. A. H., Hagtvet, B. E., Hulme, C., & Melby-Lervåg, M. (2019). Pathways to reading comprehension: A longitudinal study from 4 to 9 years of age. *Journal of Educational Psychology*, *111*(5), 751–763. http://dx.doi.org/10.1037/edu0000321

Holler, J., & Levinson, S. C. (2019). Multimodal language processing in human communication. *Trends in Cognitive Sciences*, *23*(8), 639–652. https://doi.org/10.1016/j.tics.2019.05.006

Holyoak, K. J., & Stamenković, D. (2018). Metaphor comprehension: A critical review of theories and evidence. *Psychological Bulletin*, *144*(6), 641. https://psycnet.apa.org/doi/10.1037/bul0000145

Horvath, S., Kueser, J. B., Kelly, J., & Borovsky, A. (2022). Difference or delay? Syntax, semantics, and verb vocabulary development in typically developing and late-talking toddlers. *Language Learning and Development, 18*(3), 352–376. https://doi.org/10.1080/15475441.2021.1977645

Hostetter, A. B., & Alibali, M. W. (2019). Gesture as simulated action: Revisiting the framework. *Psychonomic Bulletin & Review, 26*, 721–752. https://doi.org/10.3758/s13423-018-1548-0

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review, 15*, 495–514. https://doi.org/10.3758/PBR.15.3.495

Hoyos, C., Horton, W. S., Simms, N. K., & Gentner, D. (2020). Analogical comparison promotes theory-of-mind development. *Cognitive Science, 44*(9), e12891. https://doi.org/10.1111/cogs.12891

Huang, S. F., Oi, M., & Taguchi, A. (2015). Comprehension of figurative language in Taiwanese children with autism: The role of theory of mind and receptive vocabulary. *Clinical Linguistics & Phonetics, 29*(8–10), 764–775. https://doi.org/10.3109/02699206.2015.1027833

Huebner, B. (2013). Socially embedded cognition. *Cognitive Systems Research, 25*. 13–18. https://doi.org/10.1016/j. cogsys.2013.03.006

Ibáñez, A., Manes, F., Escobar, J., Trujillo, N., Andreucci, P., & Hurtado, E. (2010). Gesture influences the processing of figurative language in non-native speakers: ERP evidence. *Neuroscience Letters*, *471*(1), 48–52. https://doi.org/10.1016/j.neulet.2010.01.009

Iverson, J. M., Northrup, J. B., Leezenbaum, N. B., Parladé, M. V., Koterba, E. A., & West,

K. L. (2018). Early gesture and vocabulary development in infant siblings of children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *48*(1), 55–71. https://doi.org/10.1007/s10803-017-3297-8

Jafari, M., & Ansari-Pour, N. (2019). Why, when and how to adjust your P values?. *Cell Journal (Yakhteh), 20(*4), 604–607. https://doi.org/10.22074/cellj.2019.5992

Joue, G., Boven, L., Willmes, K., Evola, V., Demenescu, L. R., Hassemer, J., Hassemer, J., Mittelberg, I., Mathiak, K., Schneider, F., & Habel, U. (2020). Metaphor processing is supramodal semantic processing: The role of the bilateral lateral temporal regions in multimodal communication. *Brain and Language*, *205*, 104772. https://doi.org/10.1016/j.bandl.2020.104772

Kalandadze, T., Norbury, C. F., Nærland, T., & Næss, K. A. B. (2018). Figurative language comprehension in individuals with autism spectrum disorder: A meta-analytic review. *Autism, 22*(2), 99–117. https://doi.org/10.1177/1362361316668652

Kasirer, A., & Mashal, N. (2017). Comprehension and generation of metaphoric language in children, adolescents, and adults with dyslexia. *Dyslexia, 23*(2), 99–118. https://doi.org/10.1002/dys.1550

Kauschke, C., and Siegmüller, J. (2010). *Patholinguistische Diagnostik bei Sprachentwicklungsstörungen (PDSS)* [Diagnosis of Developmental Language Disorders], 2nd Edition. Urban & Fischer.

Kelly, S. D., Özyürek, A., & Maris, E. (2010). Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science*, *21*(2), 260–267. https://doi.org/10.1177/0956797609357327

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press.

Khatin-Zadeh, O. (2023). Embodied Metaphor Processing: A Study of the Priming Impact of Congruent and Opposite Gestural Representations of Metaphor Schema on Metaphor

Comprehension. *Metaphor and Symbol, 38*(1), 70–80. https://doi.org/10.1080/10926488.2022.2122830

Khatin-Zadeh, O., Farsani, D., Hu, J., Eskandari, Z., Zhu, Y., & Banaruee, H. (2023). A review of studies supporting metaphorical embodiment. *Behavioral Sciences, 13*(7), 585. https://doi.org/10.3390/bs13070585

Khatin-Zadeh, O., Hu, J., Marmolejo-Ramos, F., & Farsani, D. (2023). The impact of gestural representation of metaphor schema on metaphor comprehension. *Poznan Studies in Contemporary Linguistics, 59*(1), 117–131. https://doi.org/10.1515/psicl-2022-1056

Kim, M., & Schachner, A. (2020). Gesture Production and Theory of Mind: Effective Disambiguation in Communication through Gesture. *Proceedings of the Annual Meeting of the Cognitive Science Society, 42*. https://escholarship.org/uc/item/0n43m5wc

Kim, Y. S. G., Dore, R., Cho, M., Golinkoff, R., & Amendum, S. J. (2021). Theory of mind, mental state talk, and discourse comprehension: Theory of mind process is more important for narrative comprehension than for informational text comprehension. *Journal of Experimental Child Psychology, 209*, 105181, 1–20. https://doi.org/10.1016/j.jecp.2021.105181

Kirk, E., Donnelly, S., Furman, R., Warmington, M., Glanville, J., & Eggleston, A. (2022). The relationship between infant pointing and language development: A meta-analytic review. *Developmental Review, 64*, 101023. https://doi.org/10.1016/j.dr.2022.101023

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, *48*(1), 16–32. https://doi.org/10.1016/S0749-596X(02)00505-3

Lagattuta, K. H., & Kramer, H. J. (2021). Advanced theory of mind in middle childhood and adulthood. In R. T. Devine & S. Lecce (Eds.), *Theory of Mind in Middle Childhood and Adolescence* (pp. 15–36). Routledge. https://doi.org/10.4324/9780429326899

Lakoff, G. & Johnson, M. (1980). *Metaphors we live by.* University of Chicago Press.

Landerl, K., Wimmer, H., & Frith, U. (1997). The impact of orthographic consistency on dyslexia: A German-English comparison. *Cognition, 63*(3), 315–334. https://doi.org/10.1016/S0010-0277(97)00005-X

Language and Reading Research Consortium, & Chiu, Y. D. (2018). The simple view of reading across development: Prediction of grade 3 reading comprehension from prekindergarten skills. *Remedial and Special Education*, *39*(5), 289–303. https://doi.org/10.1177/0741932518762055

Lapaire, J. R. (2011). Grammar, gesture and cognition: insights from multimodal utterances and applications for gesture analysis. *Visnyk of the Lviv University. Series Philology, 102*, 88–103. https://shs.hal.science/halshs-01629052/

Lecce, S., Ronchi, L., Del Sette, P., Bischetti, L., & Bambini, V. (2019). Interpreting physical and mental metaphors: Is Theory of Mind associated with pragmatics in middle childhood? *Journal of Child Language, 46*(2), 393–407. https://doi.org/10.1017/S030500091800048X

Levorato, M. C., Nesi, B., & Cacciari, C. (2004). Reading comprehension and understanding idiomatic expressions: A developmental study. *Brain and Language*, *91*(3), 303–314. https://doi.org/10.1016/j.bandl.2004.04.002

Liszkowski, U., Brown, P., Callaghan, T., Takada, A., & De Vos, C. (2012). A prelinguistic gestural universal of human communication. *Cognitive Science, 36*(4), 698–713. https://doi.org/10.1111/j.1551-6709.2011.01228.x

Liszkowski, U., Carpenter, M., & Tomasello, M. (2008). Twelve-month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *Cognition*, *108*(3), 732–739. https://doi.org/10.1016/j.cognition.2008.06.013

Liszkowski, U., Carpenter, M., & Tomasello, M. (2007). Reference and attitude in infant pointing. *Journal of Child Language, 34*(1), 1–20.

https://doi.org/10.1017/S0305000906007689

Liszkowski, U., & Tomasello, M. (2011). Individual differences in social, cognitive, and morphological aspects of infant pointing. *Cognitive Development*, *26*(1), 16–29. https://doi.org/10.1016/j.cogdev.2010.10.001

Lorusso, M. L., Burigo, M., Borsa, V., & Molteni, M. (2015). Processing sentences with literal versus figurative use of verbs: an ERP study with children with language impairments, nonverbal impairments, and typical development. *Behavioural Neurology*, *2015,* 475271, 1–21. https://doi.org/10.1155/2015/475271

Lorusso, M. L., Galli, R., Libera, L., Gagliardi, C., Borgatti, R., & Hollebrandse, B. (2007). Indicators of theory of mind in narrative production: a comparison between individuals with genetic syndromes and typically developing children. *Clinical Linguistics & Phonetics, 21*(1), 37–53. https://doi.org/10.1080/02699200600565871

Lüke, C. (2015). Gestische Kommunikation als Vorläufer von Sprache [Gestural Communication as a Precursor to Language]. Peter Lang.

Lüke, C., Grimminger, A., Rohlfing, K. J., Liszkowski, U., & Ritterfeld, U. (2017). In infants' hands: Identification of preverbal infants at risk for primary language delay. *Child Development, 88*(2), 484–492. https://doi.org/10.1111/cdev.12610

Lüke, C., Kauschke, C., Dohmen, A., Haid, A., Leitinger, C., Männel, C., Penz, T., Sachse, S., Scharff Rethfeldt, W., Spranger, J., Vogt, S., Niederberger, M., & Neumann, K. (2023). Definition and terminology of developmental language disorders—Interdisciplinary consensus across German-speaking countries. *PLoS ONE, 18*(11), e0293736. https://doi.org/10.1371/journal.pone.0293736

Lüke, C., Leinweber, J., & Ritterfeld, U. (2019). Walking, pointing, talking–the predictive value of early walking and pointing behavior for later language skills. *Journal of Child Language, 46*(6), 1228–1237. https://doi.org/10.1017/S0305000919000394

Lüke, C., Ritterfeld, U., Grimminger, A., Liszkowski, U., & Rohlfing, K. J. (2017).

Development of pointing gestures in children with typical and delayed language acquisition. *Journal of Speech, Language, and Hearing Research*, *60*(11), 3185–3197. https://doi.org/10.1044/2017_JSLHR-L-16-0129

Lüke, C., Ritterfeld, U., Grimminger, A., Rohlfing, K. J., & Liszkowski, U. (2020). Integrated Communication System: Gesture and Language Acquisition in Typically Developing Children and Children with LD and DLD. *Frontiers in Psychology*, *11*, 118. https://doi.org/10.3389/fpsyg.2020.00118

Mainela-Arnold, E., Alibali, M. W., Hostetter, A. B., & Evans, J. L. (2014). Gesture–speech integration in children with specific language impairment. *International Journal of Language & Communication Disorders*, *49*(6), 761–770. https://doi.org/10.1111/1460-6984.12115

Manwaring, S. S., Swineford, L., Mead, D. L., Yeh, C.-C., Zhang, Y., & Thurm, A. (2019). The gesture–language association over time in toddlers with and without language delays. *Autism & Developmental Language Impairments, 4*. https://doi.org/10.1177/2396941519845545

McGillion, M., Herbert, J. S., Pine, J., Vihman, M., DePaolis, R., Keren-Portnoy, T., & Matthews, D. (2017). What paves the way to conventional language? The predictive value of babble, pointing, and socioeconomic status. *Child Development*, *88*(1), 156–166. https://doi.org/10.1111/cdev.12671

McNeill, D. (2014). The emblem as metaphor. In M. Seyfeddinipur & M. Gullberg (Eds.), *From gesture in conversation to visible action as utterance* (pp. 75–94). John Benjamins.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press.

McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review, 92*(3), 350–371. https://doi.org/10.1037/0033-295X.92.3.350

McNeill, D., & Levy, E. (1982). Conceptual representations in language activity and gesture.

In R. J. Jarvella & W. Klein (Eds.), *Speech, place, and action* (pp. 271–295). Wiley.

Miller, C. A. (2006). Developmental relationships between language and theory of mind. *American Journal of Speech-Language Pathology, 15*, 142–154. https://doi.org/10.1044/1058-0360(2006/014)

Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Development*, *78*(2), 622–646. https://doi.org/10.1111/j.1467-8624.2007.01018.x

Nilsson, K. K., & de Lopez, K. J. (2016). Theory of mind in children with specific language impairment: A systematic review and meta-analysis. *Child Development*, *87*(1), 143–153. https://doi.org/10.1111/cdev.12462

Norbury, C. F. (2005). The relationship between theory of mind and metaphor: Evidence from children with language impairment and autistic spectrum disorder. *British Journal of Developmental Psychology*, *23*(3), 383–399. https://doi.org/10.1348/026151005X26732

Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vamvakas, G., & Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: Evidence from a population study. *Journal of Child Psychology and Psychiatry, 57*(11), 1247–1257. https://doi.org/10.1111/jcpp.12573

Novack, M. A., & Goldin-Meadow, S. (2017). Gesture as representational action: A paper about function. *Psychonomic Bulletin & Review, 24*, 652–665. https://doi.org/10.3758/s13423-016-1145-z

Oesch, N., & Dunbar, R. I. (2017). The emergence of recursion in human language: mentalising predicts recursive syntax task performance. *Journal of Neurolinguistics, 43*, 95–106. https://doi.org/10.1016/j.jneuroling.2016.09.008

Osterhaus, C., & Koerber, S. (2021a). Social cognition during and after kindergarten: The relations between first-order and advanced theories of mind. *European Journal of Developmental Psychology, 18*(4), 573–592.

https://doi.org/10.1080/17405629.2020.1820861

Osterhaus, C., & Koerber, S. (2021b). The development of advanced theory of mind in middle childhood: A longitudinal study from age 5 to 10 years. *Child Development, 92*(5), 1872–1888. https://doi.org/10.1111/cdev.13627

Osterhaus, C., Kristen-Antonow, S., Kloo, D., & Sodian, B. (2022). Advanced scaling and modeling of children’s theory of mind competencies: Longitudinal findings in 4- to 6-year-olds. *International Journal of Behavioral Development, 46*(3), 251–259. https://doi.org/10.1177/01650254221077334

Özçalışkan, Ş. (2005). On learning to draw the distinction between physical and metaphorical motion: Is metaphor an early emerging cognitive and linguistic capacity? *Journal of Child Language*, *32*(2), 291. https://doi.org/10.1017/S0305000905006884

Özyürek, A. (2014). Hearing and seeing meaning in speech and gesture: insights from brain and behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130296. https://doi.org/10.1098/rstb.2013.0296

Papafragou, A. (2001). Linking early linguistic and conceptual capacities: The role of theory of mind. In A. Cienki, B. Luka & M. Smith (Eds.), *Conceptual structure, discourse and language IV* (pp.169–184). University of Chicago Press.

Paul, R., & Ellis Weismer, S. (2013). Late talking in context: The clinical implications of delayed language development. In L. Rescorla & P. S. Dale (Eds.), *Late talkers: Language development, interventions, and outcomes* (pp. 203–217). Brookes.

Parrill, F., Lavanty, B., Bennett, A., Klco, A., & Demir-Lira, O. E. (2018). The relationship between character viewpoint gesture and narrative structure in children. *Language and Cognition, 10*(3), 408–434. https://doi.org/10.1017/langcog.2018.9

Pecher, D., Boot, I., & Van Dantzig, S. (2011). Abstract concepts: Sensory-motor grounding, metaphors, and beyond. In B. Ross (Ed.), *The Psychology of Learning and Motivation,* *Vol. 54* (pp. 217–248). Academic Press.

Perner, J., & Wimmer, H. (1985). “John thinks that Mary thinks that…” attribution of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology*, *39*(3), 437–471. https://doi.org/10.1016/0022-0965(85)90051-7

Perry, L. K., Kucker, S. C., Horst, J. S., & Samuelson, L. K. (2023). Later bloomer or language disorder? Differences in toddler vocabulary composition associated with long-term language outcomes. *Developmental Science, 26*(4), e13342. https://doi.org/10.1111/desc.13342

Peterson, C. C., Wellman, H. M., & Slaughter, V. P. (2012). The mind behind the message: Advancing theory-of-mind scales for typically developing children, and those with deafness, autism, or Asperger syndrome. *Child Development*, *83*(2), 469–485. https://doi.org/10.1111/j.1467-8624.2011.01728.x

Pouscoulous, N., & Tomasello, M. (2020). Early birds: Metaphor understanding in 3-year-olds. *Journal of Pragmatics*, *156*, 160–167. https://doi.org/10.1016/j.pragma.2019.05.021

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*(4), 515–526. https://dx.doi.org/10.1017/S0140525X00076512

Pronina, M., Prieto, P., Bischetti, L., & Bambini, V. (2023). Expressive pragmatics and prosody in young preschoolers are more closely related to structural language than to mentalizing. *Language Learning and Development, 19*(3), 323–344. https://doi.org/10.1080/15475441.2022.2074852

Psyridou, M., Eklund, K., Poikkeus, A. M., & Torppa, M. (2018). Reading outcomes of children with delayed early vocabulary: A follow-up from age 2–16. *Research in Developmental Disabilities*, *78*, 114–124. https://doi.org/10.1016/j.ridd.2018.05.004

Rakhlin, N., Kornilov, S. A., Reich, J., Babyonyshev, M., Koposov, R. A., & Grigorenko, E. L. (2011). The relationship between syntactic development and theory of mind: Evidence from a small-population study of a developmental language disorder. *Journal of Neurolinguistics*, *24*(4), 476–496. https://doi.org/10.1016/j.jneuroling.2011.03.001

Rakoczy, H., Wandt, R., Thomas, S., Nowak, J., & Kunzmann, U. (2018). Theory of mind and wisdom: The development of different forms of perspective-taking in late adulthood. *British Journal of Psychology*, *109*(1), 6–24. https://doi.org/10.1111/bjop.12246

Rakoczy, H., Harder-Kasten, A., & Sturm, L. (2012). The decline of theory of mind in old age is (partly) mediated by developmental changes in domain-general abilities. *British Journal of Psychology*, *103*(1), 58–72. https://doi.org/10.1111/j.2044-8295.2011.02040.x

Rescorla, L. (2013). Late-talking toddlers: A 15-year follow-up. In L. Rescorla & P. S. Dale (Eds.), *Late talkers: Language development, interventions, and outcomes* (pp. 219–239). Brookes.

Rescorla, L. (2011). Late talkers: Do good predictors of outcome exist? *Developmental Disabilities Research Reviews, 17*(2), 141–150. https://doi.org/10.1002/ddrr.1108

Rescorla, L. (2009). Age 17 language and reading outcomes in late-talking toddlers: Support for a dimensional perspective on language delay. *Journal of Speech, Language, and Hearing Research, 52*, 16–30. https://doi.org/10.1044/1092-4388(2008/07-0171)

Rescorla, L. (2005). Age 13 language and reading outcomes in late-talking toddlers. *Journal of Speech, Language, and Hearing Research, 48*, 459–472. https://doi.org/10.1044/1092-4388(2005/031)

Rescorla, L. (2002). Language and reading outcomes to age 9 in late-talking toddlers. *Journal of Speech, Language, and Hearing Research, 45*, 360–371. https://doi.org/10.1044/1092-4388(2002/028)

Rescorla, L., Roberts, J., & Dahlsgaard, K. (1997). Late talkers at 2: Outcome at age 3. *Journal of Speech, Language, and Hearing Research, 40*, 556–566. https://doi.org/10.1044/jslhr.4003.556

Rohlfing, K. J., Lüke, C., Liszkowski, U., Ritterfeld, U., & Grimminger, A. (2022). Developmental Paths of Pointing for Various Motives in Infants with and without Language Delay. *International Journal of Environmental Research and Public Health, 19*(9), 4982. https://doi.org/10.3390/ijerph19094982

Rowe, M. L., & Goldin-Meadow, S. (2009). Differences in early gesture explain SES disparities in child vocabulary size at school entry. *Science*, *323*(5916), 951–953. https://doi.org/10.1126/science.1167025

Rundblad, G. (2017). Metaphor acquisition and use in individuals with neurodevelopmental disorders. In E. Semino & Z. Demjen (Eds.), *The Routledge Handbook of Metaphor and Language* (pp. 486–502). Routledge.

Sansavini, A., Favilla, M. E., Guasti, M. T., Marini, A., Millepiedi, S., Di Martino, M. V., Vecchi, S., Battajon, N., Bertolo, L., Capirci, O., Carretti, B., Colatei, M. P., Frioni, C., Marotta, L., Massa, S., Michelazzo, L., Pecini, C., Piazzalunga, S., Pieretti, M., Rinaldi, P., Salvadorini, R., Termine, C., Zuccarini, M., D'Amico, S., De Cagno, A. G., Levorato, M. C., Rossetto, T., & Lorusso, M. L. (2021). Developmental Language Disorder: Early Predictors, Age for the Diagnosis, and Diagnostic Tools. A Scoping Review. *Brain Sciences*, *11*(5), 654–691. https://doi.org/10.3390/brainsci11050654

Schaunig, Ines, Willinger, Ulrike, & Formann, Anton K. (2004). Das Verständnis metaphorischer Sprache bei Grundschulkindern [Children's Comprehension of Metaphoric Language]. *Zeitschrift für Pädagogische Psychologie*, 18(1), 53–61. https://doi.org/10.1024/1010-0652.18.1.53

Schneider, D., Slaughter, V. P., & Dux, P. E. (2017). Current evidence for automatic Theory of Mind processing in adults. *Cognition, 162*, 27–31. https://doi.org/10.1016/j.cognition.2017.01.018

Schnell, Z. (2007). Metaphor processing and the acquisition of idioms: A mentalistic model. *Acta Linguistica Hungarica*, *54*(1), 73–104. https://doi.org/10.1556/aling.54.2007.1.3

Smit, L., Knoors, H., Hermans, D., Verhoeven, L., & Vissers, C. (2019). The interplay between theory of mind and social emotional functioning in adolescents with communication and language problems. *Frontiers in Psychology*, *10*, 1488. https://doi.org/10.3389/fpsyg.2019.01488

Sodian, B., & Kristen-Antonow, S. (2015). Declarative Joint Attention as a Foundation of Theory of Mind. *Developmental Psychology, 51*(9), 1190–1200. https://doi.org/10.1037/dev0000039

Sodian, B., Kristen-Antonow, S., & Kloo, D. (2020). How does children's theory of mind become explicit? A review of longitudinal findings. *Child Development Perspectives, 14*(3), 171–177. https://doi.org/10.1111/cdep.12381

Spanoudis, G. (2016). Theory of mind and specific language impairment in school-age children. *Journal of Communication Disorders*, *61*, 83–96. https://doi.org/10.1016/j.jcomdis.2016.04.003

Stothard, S. E., Snowling, M. J., Bishop, D. V. M., Chipchase, B. B., & Kaplan, C. A. (1998). Language-impaired preschoolers: A follow-up into adolescence. *Journal of Speech, Language, and Hearing Research, 41*(2), 407–418. https://doi.org/10.1044/jslhr.4102.407

Sweller, N., Choi, A. J., & Austin, E. (2024). Gesture production at encoding supports narrative recall. *Psychological Research, 88*(2), 535–546. https://doi.org/10.1007/s00426-023-01886-w

Szagun, G., Stumper, B., and Schramm, S. A. (2009). *FRAKIS. Fragebogen zur frühkindlichen Sprachentwicklung* [Questionnaire regarding early child language development]. Pearson.

Tager-Flusberg, H., & Joseph, R. M. (2005). How language facilitates the acquisition of false belief understanding in children with autism. In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 298–319). Oxford University Press. https://psycnet.apa.org/doi/10.1093/acprof:oso/9780195159912.003.0014

Tellegen, P. J., Laros, J. A., and Petermann, F. (2007). *SON-R 2 1/2–7: Non-Verbaler Intelligenztest [Non-verbal intelligence test]*. Hogrefe.

Thal, D., Marchman, V. A., & Tomblin, J. B. (2013). Late-talking toddlers: Characterization and prediction of continued delay. In L. Rescorla & P. S. Dale (Eds.), *Late talkers: Language development, interventions, and outcomes* (pp. 169–201). Brookes.

Tolksdorf, N. F., Viertel, F. E., Crawshaw, C. E., & Rohlfing, K. J. (2021, June 24–30). Do shy children keep more distance from a social robot? Exploring shy children's proxemics with a social robot or a human [Paper presentation]. In J. A. Fails & M. Landoni (Eds.), *Interaction Design and Children (IDC '21, Athens, Greece)* (pp.527–531). Association for Computing Machinery. https://doi.org/10.1145/3459990.3465181

Tolksdorf, N. F., Crawshaw, C. E., & Rohlfing, K. J. (2021). Comparing the Effects of a Different Social Partner (Social Robot vs. Human) on Children's Social Referencing in Interaction. *Frontiers in Education, 5:569615*, 1–12. https://doi.org/10.3389/feduc.2020.569615

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

Tonini, E., Bischetti, L., Del Sette, P., Tosi, E., Lecce, S., & Bambini, V. (2023). The relationship between metaphor skills and Theory of Mind in middle childhood: Task and developmental effects. *Cognition, 238*, 105504. https://doi.org/10.1016/j.cognition.2023.105504

Valle, A., Massaro, D., Castelli, I., & Marchetti, A. (2015). Theory of mind development in adolescence and early adulthood: The growing complexity of recursive thinking ability. *Europe's Journal of Psychology, 11*(1), 112–124. https://doi.org/10.5964/ejop.v11i1.829

Veneziano, E. (2009). Language and internal states: A long developmental history at different levels of functioning. *Rivista Di Psicolinguistica Applicata, 9*(3), 15–34.

Vissers, C., & Koolen, S. (2016). Theory of mind deficits and social emotional functioning in preschoolers with specific language impairment. *Frontiers in Psychology, 7*, 1–7. https://doi.org/10.3389/fpsyg.2016.01734

Vogt, S., & Indefrey, P. (2017). Metaphernerwerb: eine empirische Studie bei Kindern im Alter von sechs bis vierzehn Jahren [Metaphor acquisition: an empirical study with children between the ages of six to fourteen years]. *Metaphorik.de*, 27/2017, 69–106. https://www.metaphorik.de/de/journal/27/metaphernerwerb-eine-empirische-studie-bei-kindern-im-alter-von-sechs-bis-vierzehn-jahren.html

Vulchanova, M., Milburn, E., Vulchanov, V., & Baggio, G. (2019). Boon or burden? The role of compositional meaning in figurative language processing and acquisition. *Journal of Logic, Language and Information*, *28*(2), 359–387. https://doi.org/10.1007/s10849-019-09282-7

Vulchanova, M., Saldaña, D., Chahboun, S., & Vulchanov, V. (2015). Figurative language processing in atypical populations: the ASD perspective. *Frontiers in Human Neuroscience*, *9*, 24. https://doi.org/10.3389/fnhum.2015.00024

Warnell, K. R., & Redcay, E. (2021). Neural bases of theory of mind in middle childhood and adolescence. In R. T. Devine & S. Lecce (Eds.), *Theory of Mind in Middle Childhood and Adolescence* (pp. 77–97). Routledge. https://doi.org/10.4324/9780429326899

Watson, L. R., Crais, E. R., Baranek, G. T., Dykstra, J. R., & Wilson, K. P. (2013). Communicative gesture use in infants with and without autism: A retrospective home video study. *American Journal of Speech-Language Pathology, 22*(1), 25–39. https://doi.org/10.1044/1058-0360(2012/11-0145)

Weimer, A. A., Warnell, K. R., Ettekal, I., Cartwright, K. B., Guajardo, N. R., & Liew, J. (2021). Correlates and antecedents of theory of mind development during middle childhood and adolescence: An integrated model. *Developmental Review, 59*, 100945.

https://doi.org/10.1016/j.dr.2020.100945

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, *72*(3), 655–684. https://doi.org/10.1111/1467-8624.00304

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, *75*(2), 523–541. https://doi.org/10.1111/j.1467-8624.2004.00691.x

Wesp, R., Hesse, J., Keutmann, D., & Wheaton, K. (2001). Gestures maintain spatial imagery. *American Journal of Psychology, 114*(4), 591–600. https://doi.org/10.2307/1423612

White, S., Hill, E., Happé, F., & Frith, U. (2009). Revisiting the strange stories: Revealing mentalizing impairments in autism. *Child Development*, *80*(4), 1097–1117. https://doi.org/10.1111/j.1467-8624.2009.01319.x

Wimmer, H., & Mayringer, H. (2014). *Salzburger Lese-Screening für die Schulstufen 2–9: SLS 2–9* [Salzburg Reading-Screening-Test for School Levels 2–9: SLS 2–9]. Huber.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103–128. https://doi.org/10.1016/0010-0277(83)90004-5

Wray, C., Norbury, C. F., & Alcock, K. (2016). Gestural abilities of children with specific language impairment. *International Journal of Language & Communication Disorders*, *51*(2), 174–182. https://doi.org/10.1111/1460-6984.12196

Wray, C., Saunders, N., McGuire, R., Cousins, G., & Norbury, C. F. (2017). Gesture production in language impairment: it's quality, not quantity, that matters. *Journal of Speech, Language, and Hearing Research, 60*(4), 969–982. https://doi.org/10.1044/2016_JSLHR-L-16-0141

Zdrazilova, L., Sidhu, D. M., & Pexman, P. M. (2018). Communicating abstract meaning:

Concepts revealed in words and gestures. *Philosophical Transactions of the Royal Society B: Biological Sciences, 373*(1752), 20170138. https://doi.org/10.1098/rstb.2017.0138

Zhao, X., Lynch, J. G., Jr., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and Truths about Mediation Analysis. *Journal of Consumer Research, 37*(2), 197–206. https://doi.org/10.1086/651257

Zubrick, S. R., Taylor, C. L., Rice, M. L., & Slegers, D. W. (2007). Late language emergence at 24 months: An epidemiological study of prevalence, predictors, and covariates. *Journal of Speech, Language, and Hearing Research, 50*, 1562–1592. https://doi.org/10.1044/1092-4388(2007/106)

Appendix A: Publication I

Crawshaw, C. E., Lüke, C., & Ritterfeld, U. (2025). Does early gesture usage contribute alongside oral language to later theory of mind performance and metaphor comprehension? Indications from a longitudinal study with children aged 1–9 years. *Language Acquisition*, 1–31. https://doi.org/10.1080/10489223.2025.2455174

This article was published under a CC BY-NC-ND Open Access licence and is therefore provided here in full without any changes.

LANGUAGE ACQUISITION
https://doi.org/10.1080/10489223.2025.2455174

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Does early gesture usage contribute alongside oral language to later theory of mind performance and metaphor comprehension? Indications from a longitudinal study with children aged 1–9 years

Camilla E. Crawshaw[a], Carina Lüke[b], and Ute Ritterfeld[a]

[a]TU Dortmund University; [b]Julius-Maximilians-University of Würzburg

ABSTRACT

Longitudinal work has demonstrated both that early gestural skills predict the course of children's language development and, separately, that language skills are important for the development of theory of mind (ToM) and metaphor comprehension. To extend and connect these findings, the current study explores the contributions of gesture (pointing at 1;0 and iconic gesture comprehension at 3;0) and language skills at 3;0 (productive vocabulary, productive and receptive grammar) to the development of later, more complex social pragmatic and socio-cognitive skills at 9;0: ToM and metaphor comprehension. Participants were 35 (18 boys) monolingual German-speaking children who took part within a longitudinal study between the ages of 1-9 years. Results showed that index-finger pointing at 1;0 significantly predicted metaphor comprehension at 9;0 and that both iconic gesture comprehension and language skills at 3;0 mediated the relation between index-finger pointing at 1;0 and ToM performance and metaphor comprehension at 9;0. These findings suggest an early facilitative role of gestures in the ongoing development of children's socio-cognitive and social pragmatic capacities alongside their language skills. Future studies on ToM or metaphor comprehension should thus consider and integrate gesture within their paradigms. Intervention studies with individuals who typically struggle with ToM and other socio-cognitive or social pragmatic skills may also benefit from exploring gesture within their specific contexts.

ARTICLE HISTORY
Received 9 December 2022
Accepted 13 January 2025

1. Introduction and prior work

Gesture and spoken language form an integrated communication system (e.g., Kita & Özyürek 2003). Gesture could be described as a form of "embodied" language, partially predating the ability to verbalize, and may thus contribute to embodied social cognition. These cognitive processes may involve both early theory of mind (ToM; e.g., Fenici 2012) and metaphor processing (Khatin-Zadeh, Farsani, et al. 2023), with later ToM abilities possibly embedded in external linguistic practice (Fenici 2012). The research literature spanning the development of language, gesture, ToM, and metaphor comprehension is huge in scope and vast in volume. However, the interdependencies of these skills are still not fully understood. Our focus here is to highlight the role of early gesture. We will attempt to provide a starting point for connecting its co-development alongside spoken language with the later development of metaphor comprehension and ToM, within the context of a longitudinal perspective covering ages 1–9 years. The following introductory sections are guided by the question: How might

CONTACT Camilla E. Crawshaw camilla.crawshaw@tu-dortmund.de TU Dortmund University, Rehabilitationswissenschaften, Sprache und Kommunikation, Emil-Figge-Straße 50, Dortmund 44227, Germany.

emerging co-speech gesture alongside spoken language support the later development of more complex, language-related, socio-cognitive and social pragmatic abilities such as ToM and metaphor comprehension? Given the extent of existing work addressing each of these competencies separately, we will briefly introduce each one but will then limit our scope to their connections. To help structure our review of the literature and to offer a point of reference for the reader, we have created a visual representation (Figure 1). This figure identifies relevant theoretical themes within each competency (spoken language, gesture, metaphor, and ToM). It also notes any prior work that has addressed two or more of these competencies in relation to one another. Figure 1 serves as a base from which we gradually begin to expand the perspective of our literature review to progressively incorporate further competencies and take on a more holistic approach.

The areas we address here are gesture, spoken language, metaphor (comprehension), and ToM but these skills emerge and advance at different points in a child's development, multifariously supporting and interacting with one another. There has long been a debate within the literature about whether language or ToM predate one another, although it is generally accepted that the interaction between them across development is bidirectional (e.g., de Villiers 2007). Some gestural skills such as referential (deictic) pointing emerge preverbally (e.g., Liszkowski et al. 2012). However, representational (iconic) gesture emerges largely alongside spoken language (e.g., Guidetti & Nicoladis 2008). An example of this gesture type would be extending and snapping together the index and middle fingers while holding the third and fourth fingers down with the thumb to represent scissors. Moreover, metaphor comprehension skills—a form of complex language drawing on both pragmatic and socio-cognitive skills—emerge later in childhood than general spoken language or ToM (mentalizing) skills (e.g., Di Paola et al. 2020). The development of all these skills should consequently be considered within, and as part of a wider, general communication framework. This framework may also overlap theoretical

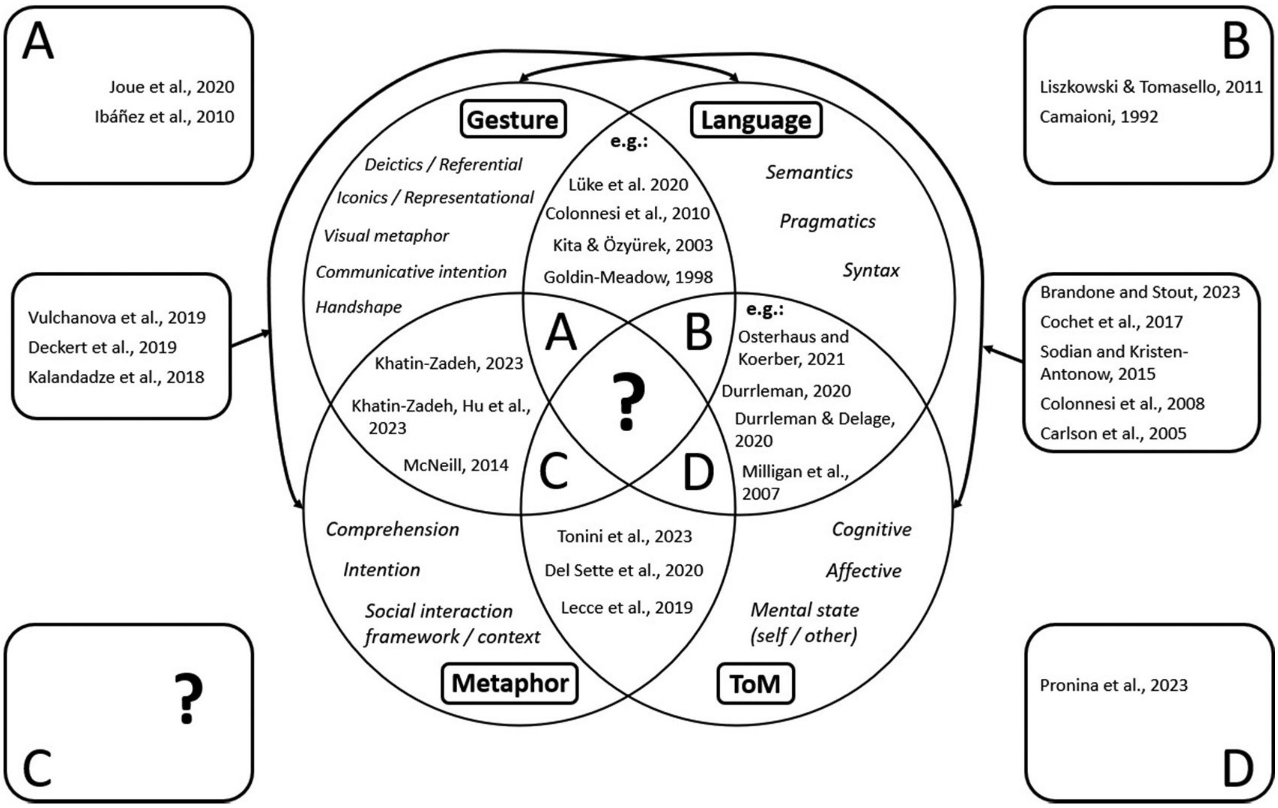


Figure 1. A Venn diagram visual representation of relevant themes that have been explored in the areas of gesture, spoken language, metaphor, and theory of mind (ToM) as well as prior work that has considered where these skills might overlap (A, B, C, D, central ?; due to lack of space, A, B, C, and D have been represented in boxes surrounding the diagram; arrows are used where only the two opposite skills have been addressed in prior work; "e.g." is used where a few examples have been selected from a wider body of work; "?" is used where no existing work could be found).

paradigms of embodiment and embeddedness (Fenici 2012). These skills are all associated with one another but metaphor comprehension appears to straddle general language (including gesture) as well as complex pragmatic and socio-cognitive ToM skills. We therefore expect additional value in considering them together instead of focusing on the core domains of metaphor comprehension and ToM in isolation from each other.

1.1. Gesture and language

Language as a whole, consists of much more than solely verbal speech; it is also nonverbal and multimodal, with gesture (nonverbal language) and speech (spoken language) forming an integrated communication system (e.g., Goldin-Meadow 1998, Graziano & Gullberg 2018, Holler & Levinson 2019, Kelly et al. 2010, Kendon 2004, Kita & Özyürek 2003, Lüke et al. 2020, McNeill 1992, Özyürek 2014). Co-speech gesture is a vital component of communication, partially predating verbal speech. Some have reasoned that gestures stem from embodied simulations of motor and perceptual states occurring during speech and thought (Hostetter & Alibali 2019). This situates gesture within the paradigm of embodied cognition; the view that our perception and conceptualization of our surroundings and interactions are informed and constrained by our bodily properties. As a visible depiction or "embodiment" of language, the construct of gesture could thus provide interesting insights into other areas of embodied social cognition.

Two types of gestures have been found to mediate both typically and atypically developing children's developmental trajectories in early language acquisition: deictic gestures (i.e., pointing) and iconic gestures (Botting et al. 2010, Colonnesi et al. 2010, Goldin-Meadow 2020, Iverson et al. 2018, Lüke et al. 2020, Mainela-Arnold et al. 2014, Rowe & Goldin-Meadow 2009, Wray et al. 2016). Although "language" within an integrated communication system includes gesture, prior research considering language skills relative to other competencies has predominantly bypassed gesture (or considered it separately), focusing solely on spoken language. Going forward we will thus refer to the two separately as "language" and "gesture."

1.2. Theory of Mind (ToM)

ToM refers to an understanding of and ability to attribute one's own and others' mental states and representations. It taps many different skills, including cognitive and affective perspective-taking, and is closely associated and intertwined with the development of linguistic, social pragmatic, and other cognitive skills (c.f. Devine & Lecce 2021). Interestingly, Fenici (2012) proposed that ToM straddles the cognitive paradigms of embodiment and embeddedness. Embedded cognition theorizes that we scaffold and enhance our understanding of external contexts and environments by cognitively embedding internal representations of ourselves within them (e.g., Huebner 2013). To cognitively embed ourselves within situations, we must first have embodied representations of ourselves. Fenici (2012) suggested that early ToM processes are rooted in the integration of sensory-motor information (embodiment) and later ToM processes are supported and scaffolded by the external structure of language (embeddedness). A key point is that ToM abilities do not emerge in isolation and they have been particularly closely linked to language development. Both primary and more complex ToM skills appear to be at least initially mediated by language abilities. Many researchers propose that language plays a causal role in ToM development although this relation is bidirectional (for a meta-analysis, see, Milligan et al. 2007).

1.3. ToM and language

Prior to addressing the specific language skills implicated in ToM development, it is important to consider what types of ToM tasks are used in research as this may affect outcomes. We will therefore first describe two of the most convincing tasks. Typically, the tasks and contexts researchers use to tap

ToM ability are presented through the medium of language, which can pose differential linguistic challenges for children. The most widely used and broadly accepted test of ToM in pre-school to early school age is the false belief (FB) task (for a recent overview, see: Devine & Lecce 2021). There are many different variations on the FB task (cf. Devine & Lecce 2021). One of those most often utilized requires children to infer a character's mistaken belief about the location of an object to explain or predict their behavior (Wimmer & Perner 1983). Children typically pass this task from between ages 4-6 years and it reliably predicts developmental changes in their mental state conceptualization and understanding during early childhood (cf. meta-analysis by Wellman et al. 2001). Variants exist to draw on more complex FB abilities. These are typically passed later, emerging at 6 years of age but with some unreliable performance up until at least the age of 9, if not further into adolescence (cf. Devine & Lecce 2021, Perner & Wimmer 1985). More advanced ToM can be measured using the Strange Stories task originally developed by Happé (1994; updated in White et al. 2009; for a discussion of other tasks: cf. Devine & Lecce 2021). It covers a wide range of more complex mental state themes, such as double-bluff, persuasion, white lie, and misunderstanding. Accurate performance on this task emerges between the ages of 6–10 years (Happé 1994, White et al. 2009). Previous work has demonstrated that it can be used as a valid and reliable predictor of ToM performance in children in middle childhood, that is, between the ages of 7–13 (Devine & Hughes 2016). Of course, since longer texts are used in the Strange Stories task to set out contexts with increased ToM complexity, this also leads to increased demands on linguistic and processing capabilities.

Regarding specific language competencies linked to ToM performance, both semantic and syntactic skills have been identified as important predictors in prior work (c.f. Astington & Baird 2005); however, their exact contributions remain a topic of debate. Conducting a meta-analysis, Milligan et al. (2007) found that receptive vocabulary measures were more weakly related to ToM FB task performance than general language measures (a combination of semantic and syntactic items), with earlier language performance more strongly affecting later FB than the reverse. Other work has identified grammatical skills (particularly sentential complementation, e.g., De Villiers & Pyers 2002, Durrleman 2020, Durrleman & Delage 2020) to be associated with FB performance. Osterhaus & Koerber (2021) identified that language ability at age 6 (assessed using a receptive grammar subtest) significantly correlated with performance on various ToM tasks (including Happé's Strange Stories and an array of FB tasks) between ages 5-10; verbal-processing demands did not explain away this correlation. However, general vocabulary (Happé 1995a) and semantic skills such as understanding mental state vocabulary have also been identified in ToM development (cf. Astington & Baird 2005, Doherty 2008). Still, the precise role played by language in ToM development is not yet well defined. This may be because it is difficult to disentangle specific components of language, since they all interact with one another across acquisition. Findings thus far might indicate that different types of ToM tasks could be associated with different aspects of language skills with relations fluctuating over time.

1.4. ToM and gesture

Gesture is tightly interwoven with spoken language and functions as another form of communication. Since language and ToM have been shown to be closely related, logically we might expect to see a similar relation between ToM and gesture. Overall language outcomes have been associated with children's understanding of intentional communication (Camaioni 1992) as well as an early use of index-finger pointing, a precursor to communicative conventionalization (Liszkowski & Tomasello 2011, Lüke et al. 2020). Pointing with the index finger differs qualitatively from whole-hand pointing. Index-finger pointing embodies children's early socio-cognitive understanding of communicative intentions and shared intentionality, offering a nonverbal method for forming joint intentions and establishing joint attention (Liszkowski & Tomasello 2011). Prior work has identified two parallel developmental progressions in pointing behavior. First, children progress from whole-hand to index-finger pointing between the ages of 12-21 months (Lüke, Ritterfeld, et al. 2017). Secondly, children progress in both the production and understanding of communicative motivations for pointing, from imperative to declarative-expressive to declarative-informative (Camaioni et al. 2004, Lüke, Grimminger, et al. 2017, Rohlfing et al. 2022). Both a point's handshape (1) and its communicative motive (2) may be relevant for developing ToM. Previous research has indicated that the production of declarative pointing is also linked to understanding others' intentions (Camaioni et al. 2004, Liszkowski & Tomasello 2011).

Some researchers have attempted to link the communicative motives for pointing gestures to the development of ToM, with mixed results. Colonnesi et al. (2008) investigated whether pointing gestures and intentional understanding abilities at 12 and 15 months predicted later comprehension of perception and intention as well as the ability to explain others' actions in a psychological way at 39 months. They found that only 12-month-olds' comprehension of imperative pointing played a role in predicting their later understanding of perception and their ability to meta-linguistically explain actions (Colonnesi et al. 2008). In contrast, Sodian & Kristen-Antonow (2015) found that earlier acquisition of general declarative pointing (but not imperative pointing) predicted ToM performance measured by FB understanding at age 4;2. Others found that only declarative-informative (and not declarative-expressive) pointing was related to ToM-scale measure performance in children between ages 3;1–4;0 (Cochet et al. 2017). The contrastive existing findings regarding associations between ToM and communicative motives for pointing may indicate that pointing in general is important for developing ToM, rather than the specific motive. Since the handshape of the point (i.e., index-finger) has also been found to be more important than the motive in predicting spoken language ability (Lüke, Grimminger, et al. 2017), it might thus be a more relevant avenue for investigation in reference to ToM. Further research is evidently necessary to establish whether the specific communicative motives for pointing gestures or their specific hand shapes predict ToM development. Brandone & Stout (2023) found that children between ages 8-12 months who more frequently attempted to initiate joint attention—measured by use of shows, points, and gaze alternation—performed better on a parent-report survey measure of mental state understanding (ToM) between ages 37-45 months. This finding might support the idea that more general pointing—rather than specific communicative motives—facilitates a pathway from early infant social cognition to later mental state reasoning. It might also fit with the idea that early ToM processes are embodied while later ToM processes are embedded in language (Fenici 2012).

Beyond the role of pointing in the development of ToM, very few studies have considered the impact of other types of gestures or a later stage of gesture development. This is surprising given that ToM entails (meta-)representation and iconic gestures are considered to be representational. To our knowledge, only one study has considered iconic gesture in relation to ToM development. Carlson et al. (2005) compared children's performance on standard verbal FB tasks (the unexpected contents and unexpected location tasks) vs. a newly designed gesture-based version, in which the experimenter and participants used gesture labels for the objects. The authors identified a transitional period in children's performance at around age 4 during which a number of children successfully achieved the gesture-based FB task but were not yet able to pass the standard verbal version (Carlson et al. 2005). As noted by Carlson et al. (2005), gesture has a number of features relevant to ToM development: (1) Gestures can convey information that is not present in accompanying speech; (2) gesture can support cognition by providing an alternative route for representing modes of thought; (3) iconic gestures are not arbitrary symbols, they present a visuospatial format of their verbal counterparts, reducing cognitive load and enriching semantic encoding; and (4) gesture-speech mismatch is a marker of cognitive transition and conceptual development in children with implications for problem-solving tasks. Like spoken language, gestures can also represent some other entity, object or abstract concept, and their visuospatial format may be more accessible to children. Pictorial and linguistic representations vs. ToM or perspective-taking have sometimes been differentiated as "non-mental" and "mental" representations (Doherty 2008). Previous work has indicated that children's metalinguistic awareness —the capacity to reflect on language's representational nature—develops in close connection with their understanding of mental representations (ToM; Doherty 2000; Doherty and Perner 1998, 2020).

It is thus possible that processing gesture bridges the gap between mental and non-mental representations, requiring or promoting the development of metalinguistic awareness and ToM.

To summarize connections so far between ToM, gesture and language: (1) both deictic and iconic gesture have been found to predict or be correlated to the development of language skills (e.g., Botting et al. 2010, Colonnesi et al. 2010, Goldin-Meadow 2020, Iverson et al. 2018, Lüke et al. 2020, Mainela-Arnold et al. 2014, Rowe & Goldin-Meadow 2009, Wray et al. 2016). (2) Language skills have been associated with concurrent ToM skills but have also been found to predict them (e.g., Durrleman 2020, Durrleman & Delage 2020, Milligan et al. 2007, Osterhaus & Koerber 2021). (3) Findings considering the contribution of gesture to ToM have been mixed regarding deictic gesture (Cochet et al. 2017, Colonnesi et al. 2008, Sodian & Kristen-Antonow 2015), while only one study has considered iconic gesture in relation to ToM, observing that a gesture-based FB task transitionally supported ToM performance (Carlson et al. 2005). (4) Gesture may embody language (Hostetter & Alibali 2019) while early ToM processes may be embodied, and later ToM processes may be embedded in language (Fenici 2012). Additional, targeted research would appear necessary to further disentangle these relationships.

1.5. Metaphor comprehension

We turn finally to metaphor comprehension since it has been linked with both language and ToM skills. Social pragmatic (e.g., metaphor comprehension) and socio-cognitive (e.g., ToM) skills become increasingly important for children at school age in achieving successful communication and social interaction (cf. Devine & Lecce 2021). Dialogue becomes more complex, and children require an ever-greater understanding of ambiguous and nonliteral figurative language. This entails general language, metalinguistic and social pragmatic (e.g., metaphor comprehension) skills as well as perspective-taking skills and a developing understanding of others' minds (ToM). Acquiring an understanding of ambiguous and nonliteral figurative language occurs later than other language skills (e.g., grammar), accelerating from the first years of school but extending throughout adolescence into young adulthood (Carriedo et al. 2016, Chahboun et al. 2016, Norbury 2005, Vulchanova et al. 2015). Some work has found evidence for implicit metaphor comprehension at as early as ages 3-5 across certain dimensions of metaphors (Özçalışkan 2005, Pouscoulous & Tomasello 2020). However, most work has focused on more explicit, general metaphorical understanding, which develops significantly between ages 7-10 (e.g., Deckert et al. 2019, Özçalışkan 2005, Vogt & Indefrey 2017) and even between ages 11-21 (Carriedo et al. 2016).

Metaphors vary across several dimensions, that is, in terms of their familiarity, syntactic structure, and linguistic context (Babarczy et al. 2019). As noted by Babarczy et al. (2019), the way authors categorize their metaphors can influence their findings. Metaphor contexts and categories differ (e.g., verbal vs. visual, novel vs. conventional, perceptual vs. abstract). They require different levels of perceptual and conceptual processing or comprehension, and some may be more cognitively or culturally salient (Babarczy et al. 2019). Context, specific verbal and nonverbal cognitive abilities as well as verbal and chronological age can therefore differentially contribute to successful comprehension (Babarczy et al. 2019, Del Sette et al. 2020, Lecce et al. 2019, Tonini et al. 2023, Vogt & Indefrey 2017). Work attempting to identify how specific abilities precisely contribute to metaphor comprehension has produced varying results but language competencies and ToM are generally targeted.

1.6. Metaphor, language and gesture

Language competencies have already been identified as crucial for metaphor comprehension. In a meta-analysis, Kalandadze et al. (2018) established that language abilities were a better predictor of metaphor comprehension than chronological age. Deckert et al. (2019) also identified semantic knowledge and comprehension of word meaning as a predictor of metaphor comprehension between ages 7–10. In a narrative review, Vulchanova et al. (2019) further identified that metaphor comprehension is affected by strength of semantic relation meaning.

A minimal amount of work has however considered gesture and metaphor in relation to one another. This is surprising because iconic gestures and metaphors share many striking similarities. First, both require "as-if" processing and the understanding that something can stand for something else, functioning as a form of representational communication. Second, both are a form of augmentative communication in that they add something to the accompanying speech or language for the listener. Third, both tap a developing understanding of communicative intentions. Fourth, both gestures (e.g., emblems: Kendon 2004, McNeill 1992) and metaphors (cf. Holyoak & Stamenković 2018, Schnell 2007) can become conventionalized within speech communities. Iconic gestures can also be metaphoric in themselves (McNeill 1992), allowing a speaker to depict metaphors visuospatially instead of linguistically (i.e., a spoken or written metaphor). Some researchers even propose that emblems (highly conventionalized gestures among certain users, e.g., holding up two fingers in the 'peace' sign; Kendon 2004, McNeill 1992) are themselves metaphors (McNeill 2014). Empirical research has only started to explore these interesting parallels. Recent work by Khatin-Zadeh and colleagues found that congruent gesturally depicted metaphor schema prime metaphor comprehension as well as increase and improve response times for metaphor sensibility judgments (Khatin-Zadeh 2023, Khatin-Zadeh, Hu, et al. 2023). Khatin-Zadeh (2023) subsequently argued that a metaphor schema's gestural depiction equates to an embodied realization of that metaphor, facilitating activation of a wider neural network representing the metaphor's meaning.

Existing neuroscientific work has also addressed a relation between gesture and metaphor. Ibáñez et al. (2010) analyzed event related potentials (ERPs) using electroencephalography (EEG) and found that metaphor comprehension is sensitive to co-speech gestures: Gestures helped advanced second-language speakers to distinguish metaphorical from literal meaning. Using functional magnetic resonance imaging (fMRI), Joue et al. (2020) compared metaphors that were unimodal (expressed only in gestures) vs. multimodal (expressed in speech with accompanying gestures). Both were found to recruit general semantic processing areas of the brain (Joue et al. 2020). Taken together, these findings could support embodied metaphor processing theories (Khatin-Zadeh 2023, Khatin-Zadeh, Farsani, et al. 2023, Khatin-Zadeh, Hu, et al. 2023).

1.7. Metaphor and ToM

Thus far, work relating ToM to metaphor comprehension has mainly been a focus within autism research (e.g., Chahboun et al. 2016; Happé 1995b; Norbury 2005; Vulchanova et al. 2015, 2019). ToM skills do appear to contribute to metaphor comprehension although sparse work has considered these associations within the typically developing (TD) population. Lecce et al. (2019) found that TD 9-year-olds performed lower than older age groups on metaphors requiring inference about mental but not physical attributes, and that this was associated with lower scores on ToM tasks. They also found that 9-year-olds who performed better on ToM measures were also better at interpreting mental but not physical metaphors (Lecce et al. 2019). The authors therefore proposed that the link between metaphor and ToM is stronger when metaphorical interpretation involves mental aspects (Lecce et al. 2019). Research by Tonini et al. (2023) examined 169 8-, 9-, and 10-year-old children's understanding of metaphors and ToM (using the Strange Stories task) alongside other linguistic skills. ToM was found to support only 9-year-olds' understanding of mental, not physical, metaphors within one task requiring them to verbally explain both kinds of metaphor (Tonini et al. 2023). However, in another task drawing on referential contexts, ToM supported both and 8- and 9-year-olds' accuracy in understanding metaphors (Tonini et al. 2023). Interestingly, in their 10-year-old group, ToM effects appeared negligible across both tasks (Tonini et al. 2023). The authors concluded that ToM has a task- and item- specific role in metaphor comprehension, and that their relation changes over time, with younger children needing to draw more heavily on ToM when faced with richer metaphorical contexts (Tonini et al. 2023). It has also been suggested that longitudinal associations between ToM and metaphor comprehension are rooted in inferential skills (Del Sette et al. 2020). Deckert et al. (2019) also found that metaphor processing was linked to an understanding of facts, principles, and social situations as well as inferential verbal reasoning. The idea that ToM overlaps embodiment and embeddedness (Fenici 2012) might thus be paralleled in metaphor comprehension: Some metaphors could be embodied in sensory-motor processing and others embedded in language reasoning.

1.8. Gesture, language, ToM, and metaphor

To bring this all together, previous work has separately demonstrated that early gestural skills predict children's language development trajectories, that ToM development may be facilitated by language skills (and vice versa), and that both language competencies and ToM are associated with metaphorical comprehension (e.g., Astington & Baird 2005, Del Sette et al. 2020, de Villiers 2007, Doherty 2008, Doherty & Perner 1998, Durrleman 2020, Durrleman & Delage 2020, Happé 1995b, Lecce et al. 2019, Lüke et al. 2020, Milligan et al. 2007, Norbury 2005). Most existing work has only connected the development of two (e.g., language and gesture; language and ToM; gesture and ToM) or, at most, three of these competencies (e.g., language, metaphor, and ToM; language, gesture, and ToM; Figure 1). There is also a shortage of studies exploring all these themes within an extended longitudinal context. It is not yet clear how children's gestural development (including both intended communicative purpose and handshape) interacts with their later development of more complex language, metaphor comprehension, and ToM. Studies linking gesture and ToM development are generally limited to work on specific communicative motives for pointing (e.g., Cochet et al. 2017, Sodian & Kristen-Antonow 2015). Only very few have considered the role played by later emerging iconic gesture in ToM development (Carlson et al. 2005). Gesture and metaphor processing have been connected and situated within the framework of embodied cognition (Khatin-Zadeh 2023, Khatin-Zadeh, Hu, et al. 2023, Joue et al. 2020). However, this has not yet been addressed from a longitudinal, developmental perspective.

To effectively use gestures in communication, children need to conceptualize how their interlocutor may perceive the intended meaning of those gestures. Gesture arguably embodies language and as a form of early, visual communication, it may pose a bridge between pictorial and linguistic representations. It may also tap metalinguistic awareness, drawing on metaphoric comprehension and ToM skills. Investigating gesture in relation to the constructs of language, metaphor comprehension and ToM would thus be particularly interesting. For instance, it may function as both a potential early scaffold and subsequent interface between elements of embodied and embedded communicative cognition. The current study seeks to begin exploring the above gaps by bringing together and building upon the findings from the existing literature, linking these competencies within a wider communication framework. In doing so, we consider children's preverbal (deictic gesture) and early communication skills (iconic gesture and spoken language) in relation to the development of their later, more complex metalinguistic (metaphor comprehension), and socio-cognitive (ToM) abilities. Since gesture emerges preverbally and may potentially bridge later development in language, ToM and metaphor comprehension, we hope to identify it as a potential predictor to be investigated further in future work.

2. Current study

Since the scope of prior work on these various interrelated domains is so vast, this study is intended to be an initial exploration within a more holistic and longitudinal context. The analyses have also been carried out with a relatively small sample size of participants ($N = 35$). Our research question thus asks: Can early gestural behavior (pointing at age 1;0 both in terms of hand shape and communicative motive; iconic gesture comprehension at age 3;0) predict outcomes at age 9;0 in language (productive vocabulary, productive and receptive grammar), ToM (hidden emotions, sarcasm, explicit false belief; Happé's Strange Stories), and metaphor comprehension skills?

Prior work has identified that: (a) earlier gestural skills mediate and predict language ability (e.g., Colonnesi et al. 2010, Lüke et al. 2020, Mainela-Arnold et al. 2014, Wray et al. 2016). It has also shown that (b) various language abilities are predictive of later ToM (Durrleman 2020, Durrleman & Delage 2020, Milligan et al. 2007, Osterhaus & Koerber 2021) and strongly related to metaphor comprehension skills (for a discussion, see Pronina et al. 2023). However, it has not yet clearly ascertained (c) a direct relationship between early deictic or iconic gesture and ToM development (Carlson et al. 2005, Cochet et al. 2017, Colonnesi et al. 2008, Sodian & Kristen-Antonow 2015). Nor has it yet detected d) a clear relationship between early gestural communicative competencies and metaphor comprehension skills. This is despite prior theory and evidence paralleling the representational and intentional natures of metaphor and gesture (Camaioni et al. 2004; Doherty 2000; Doherty and Perner 1998, 2020; Khatin-Zadeh 2023, Liszkowski & Tomasello 2011, McNeill 2014).

In light of previous work, we would expect to find that both early gestural (index-finger pointing and iconic gesture comprehension) and language skills (both vocabulary and grammar) predict later outcomes in ToM and metaphor comprehension skills. Prior work has identified that the handshape of pointing (i.e., with the index-finger rather than the whole hand) is more important for predicting spoken language ability than the communicative motive for pointing (Lüke, Grimminger, et al. 2017). We would therefore expect this to be analogous for ToM and metaphor comprehension. Given existing findings (Deckert et al. 2019, Happé 1995b, Kalandadze et al. 2018, Lecce et al. 2019, Norbury 2005, Schaunig et al. 2004; Tonini et al. 2023, Vulchanova et al. 2019), we would also expect to see concurrent correlations between ToM, metaphor comprehension, and language (but particularly vocabulary as an indicator of semantic knowledge) at age 9;0. Earlier stages of this longitudinal study have shown that index-finger pointing at age 1;0 can predict language skills up until age 6;0, mediated by both iconic gesture comprehension and language skills at age 3;0 (Lüke et al. 2020). We would consequently expect to find that these mediation pathways between gesture at 1;0 and 3;0 as well as language ability at 3;0 extend to age 9;0. Since language skills are closely related to ToM and metaphor skills, we would also expect to find mediation pathways between their outcomes at 9;0 and index-finger pointing at 1;0, via either iconic gesture comprehension or language skills at 3;0. This next stage of the longitudinal project is important because children's ToM development continues into middle childhood (at 9 years of age, they are in the center of this influential period) while more complex linguistic skills such as metaphor comprehension are still emerging. Our novel approach in relating these later developing skills to very early gesture skills could tell us more about the underlying cognitive processes involved in ToM and metaphor comprehension, facilitating very early intervention for at-risk children.

3. Method

The current work continues and extends a wider longitudinal study carried out over many years for which prior results have already been published (cf. Lüke et al. 2020, 2019; Lüke, Grimminger, et al. 2017; Lüke, Ritterfeld, et al. 2017; Rohlfing et al. 2022). This part of the longitudinal study examines the development of children's later, more complex language, socio-cognitive and social pragmatic skills (at age 9;0) in relation to their earlier gestural and communicative abilities (at ages 1;0 and 3;0).

3.1. Participants

A sample of 35 monolingual German-speaking children (18 boys, 17 girls) who participated in a longitudinal study between the ages of 12 months and 6 years was available to us at age 9 for further data collection. At the age of 6;0 years, there were 41 participating children, all of whom we contacted to ask to participate in this round of the study. The families of 2 of the children did not have time to participate, the family of 1 child was unwilling for them to participate in online data collection sessions, the families of 2 children withdrew due to personal reasons, and 1 child had to be excluded from the sample due to unwillingness to engage in the tasks during the first two sessions. Participants' socio-economic status (SES) was measured based on maternal and paternal years of education as well as household income adjusted for number and age of occupants. The children came from a western, educated, industrialized, rich, democratic (WEIRD) society and their parents tended toward a rather high educational level, yet an average household income compared to families living in Germany in the same year. Many prior points of data collection took place between ages 1;0–6;0 years as part of the wider study. However, in the current work we focus on data from the ages of 1;0 (M = 1 year 0 months 5 days, SD = 12 days), 3;0 (M = 3 years 0 months 21 days, SD = 14 days), 3;6 (M = 3 years 6 months 21 days, SD = 17 days) and now 9;0 (age at session 1: M = 9 years 0 months 16 days, SD = 11 days; age at session 2: M = 9 years 0 months 24 days, SD = 12 days; age at session 3: M = 9 years 1 month 18 days, SD = 31 days).

3.2. Ethical approval

The ethical considerations of all procedures, measures, and assessment of participants were evaluated and granted approval by the Internal Review Board of TU Dortmund University prior to commencing the study. The parents or caregivers of the children who participated in the study gave informed consent and were given the opportunity to withdraw their child and the relevant data from the study at any time. The children also gave verbal assent and could withdraw from the experimental interactions at any point in time at no disadvantage to themselves or their family.

3.3. Procedure

Due to the number of measures being tested at age 9;0, the children each participated in three data collection sessions. As a result of the ongoing pandemic, the first two sessions took place digitally using a digital conferencing tool (Zoom), while the final third session took place in person in our lab. All three sessions of data collection occurred typically within 1 month. The productive vocabulary measure and ToM-scale were conducted during session 1, the metaphor comprehension and Strange Stories tasks during session 2, and the productive and receptive grammar measures during session 3. Each child always interacted with the same experimenter (one of two research assistants) across all three sessions. Only one exception occurred where a child had the first two digital sessions with one experimenter and then had the final in-person session with the other experimenter; this was unfortunately unavoidable due to a minor delay caused by the ongoing COVID-19 pandemic. In the online sessions, all children were tested at home using either a laptop or tablet that belonged to the family of the participant. Each experimental interaction was audio- and video-recorded.

3.4. Measures and stimuli

Table 1 provides an overview of the ages at which the measures were conducted during the longitudinal study.

3.4.1. Index-finger pointing at 1;0: Decorated room

A semi-natural setting was designed to assess children's pointing behavior in terms of frequency and hand shape, occurring within a room decorated with 16 interesting objects, pictures, and events. Prior work has shown this type of method to be effective at eliciting pointing gestures from children at age 1;0 (Liszkowski & Tomasello 2011). Caregivers were asked to carry their child around the room for a period of 6 minutes while looking at (but not touching) the items. Meanwhile they were video recorded by four cameras placed in each corner of the room. The video recordings were coded afterward for all occurrences of pointing. Pointing was defined as the child extending their hand and arm at least more than halfway toward a picture or object without touching or grabbing it. These points were then further coded as either index-finger points (index finger clearly extended relative to their other fingers) or whole-hand points (index finger not clearly extended relative to their other fingers). Following this procedure, children were then classified as either index-finger pointers

Table 1. Age and measures conducted in this longitudinal study.

	Age 1;0	Age 3;0	Age 3;6	Age 9;0
Index-finger pointing (decorated room)	X			
Communicative Motives for Pointing (Declarative-Expressive, Declarative-Informative, Imperative)	X			
Iconic Gesture Comprehension		X		
Nonverbal IQ (SON-R)			X	
Productive Vocabulary		X (PDSS)		X (P-ITPA)
Productive and Receptive Grammar		X (SETK 3-5)		X (P-ITPA & TROG-D)
ToM (Extended Theory-of-Mind Scale and Happé's Strange Stories)				X
Metaphor Comprehension (explanation of sentences containing metaphors)				X

SON-R, Tellegen et al. (2007); PDSS, Kauschke & Siegmüller (2010); SETK 3-5, Grimm (2001); P-ITPA, Esser et al. (2010); TROG-D, Fox (2013); Extended Theory-of-Mind Scale, originally by Wellman & Liu (2004), Peterson et al. (2012), translated by Henning et al. (2013); Happé's Strange Stories, Happé (1994), White et al. (2009), pre-tested German translations by Rakoczy et al. (2012), (2018); Ebert (2020); Metaphor comprehension task adapted from original by Vogt & Indefrey (2017).

if they had pointed with their index finger at least once, or as whole-hand pointers if they had not pointed with their index finger at all (cf. Liszkowski & Tomasello 2011). Full details of this task and development procedure are already presented in Grimminger et al. (2019) and Lüke, Grimminger, et al. (2017).

3.4.2. Communicative motives for pointing at 1;0: Declarative-expressive, declarative-informative, imperative

Each of the pointing settings were based on procedures used in existing work: imperative pointing (Camaioni et al. 2004), declarative-expressive pointing (Liszkowski et al. 2007), and declarative-informative pointing (Liszkowski et al. 2008). In each setting, the child sat in a highchair facing an experimenter on the other side of the table. Behind the experimenter was a white cloth screen (2.30 m x 2.50 m) with four windows (2 on the left and 2 on the right, each one sized 25 cm x 25 cm). Four trials were performed with different objects for each setting with a randomized order of presentation.

In the imperative pointing setting, the child was sat 65 cm away from the table while the experimenter played one-by-one with two wind-up and two musical toys. The experimenter played with a toy for 10 seconds then looked at the child and waited for 15 seconds. The experimenter then emoted positively about the toy and waited another 15 seconds before giving the toy to the child. In the declarative-expressive pointing setting, a research assistant presented four different animal hand-puppets in a randomized order through the windows in the screen behind the experimenter. Each trial consisted of two 20-second-long phases: a present phase during which the experimenter turned to the puppet and emoted positively about it, alternating their gaze twice between the puppet and child; and an absent phase during which the experimenter alternated their gaze twice between the child and the screen while smiling and emoting positively. In the declarative-informative pointing setting, there were a total of four trials during which the children encountered objects for clearing away, a pencil and paper, or a small set of jigsaw pieces. Within each trial one component or part of the items fell off the table in front of the child while it appeared that the experimenter had not noticed. This was to elicit an informative pointing gesture from the child when the experimenter then explicitly searched for the missing component or part. These 3 settings eliciting declarative-expressive, declarative-informative, and imperative pointing as well as any differences from the original studies have been elaborated in detail within prior publications (Lüke, Grimminger, et al. 2017, Rohlfing et al. 2022). Each of the settings were video recorded and the children's gestures were later counted across trials. For our analyses, we have used a mean score of the number of gestures produced by each child across all of their trials within each individual setting.

3.4.3. Iconic gesture comprehension at 3;0

Children were tested on their comprehension of iconic gestures at age 3;0. They were required to match each of 16 (plus one further practice item: hat) iconic gestures (e.g., an action demonstrating a pair of scissors) to a target corresponding picture within a set of 4 images (one set for each target gesture). The target items consisted of 8 perceptual iconic gestures (car, ice cream, comb, knife, scissors, key, bird, toothbrush) and 8 pantomime iconic gestures (ball, tree, book, elephant, window, rabbit, house, cat). During the practice trial, the experimenter explained the task by saying: "Look, I'll show you some movements and then you show me which of these pictures matches the movement. For example, show me [GESTURE]." The child then received feedback: "Yes, that is correct. This [GESTURE] matches with a hat," if they were correct, or "No. Look, this [GESTURE] matches best with a hat [points to the picture of a hat]," if they were incorrect. For the test items, the child received no feedback from the experimenter. The children were then scored as either correct or incorrect. Full details of this task and its development procedure have been presented in a prior publication (Lüke et al. 2020).

3.4.4. Nonverbal intelligence quotient (IQ) at 3;6

A measure of the children's IQ was taken at age 3 and a half using the Snijders-Oomen nonverbal IQ test (SON-R, Tellegen et al. 2007), which we will include as a potential control variable in our analyses. It covers six subtests: (1) mosaics (spatial visualization and analysis to copy patterns within a framework), (2) categories (abstract reasoning and sorting according to principles), (3) puzzles (concrete reasoning and spatial relationship recognition to recreate or form pictures), (4) analogies (abstract reasoning and deduction of principles for sorting), (5) situations (concrete reasoning and deduction to complete pictures), and (6) patterns (spatial visualization to replicate patterns).

3.4.5. Language at 3;0

Children's language competencies were measured at age 3;0 and consisted of standardized German-language tests of productive vocabulary (PDSS: Patholinguistische Diagnostik bei Sprachentwicklungsstörungen [Diagnosis of Developmental Language Disorders], Kauschke & Siegmüller 2010) as well as receptive and productive grammar (SETK 3-5: Sprachentwicklungstest für drei- bis fünfjährige Kinder [Language development test for three- to five-year-old children], Grimm 2001). Using principal component analysis, we calculated a composite measure summarizing the children's performance across these tasks. The Kaiser-Meyer-Olkin measure verified the sampling adequacy of the analysis (KMO = .724), with one factor extracted due to its eigenvalue of 2.29 being over Kaiser's criterion of 1, and this factor explained 76.2% of the variance. To enable us to calculate this composite measure of language skills at 3;0 years for every child, missing tasks were substituted with the mean value. Full details of this calculation are reported in Lüke et al. (2020). This composite score (factor loading) was used as the variable "language at 3;0" in our subsequent analyses. We used this approach so that we could work with an indicator of early, "general" language skills (both semantic and syntactic), since correlations with each of these skills and ToM have been reported in the prior literature.

3.4.6. Productive vocabulary at 9;0

To measure children's vocabulary skills at age 9;0, we used the productive vocabulary subtest from the standardized German-language version of the (Potsdam-)Illinois Test for Psycholinguistic Abilities (P-ITPA, Esser et al. 2010). This task elicits progressively more difficult target words from the child by presenting descriptors (e.g., "I'm thinking about something with feathers, what could that be?"), tapping both semantic and lexical skills.

3.4.7. Productive and receptive grammar at 9;0

To measure children's productive grammar abilities at age 9;0, we used the expressive grammar subtest from the standardized P-ITPA (Esser et al. 2010). To assess their receptive grammar skills at age 9;0, we used the standardized German-language test of grammar comprehension (TROG-D, Fox 2013). Both tasks present the child with sets of supporting images. In the expressive grammar subtest, the experimenter points to the images and the child must then verbally complete the sentences, for example, “This is big, and this is . . . [bigger].” In the receptive grammar test, the child is presented a set of 4 images, read aloud a sentence by the experimenter, and then must point to the correctly corresponding image.

3.4.8. ToM at 9;0

Two measures of children’s ToM skills—ToM-scale and Strange Stories—were taken and analyzed separately. For the ToM-scale, we used a combined score of performance across three subtests (hidden emotion; sarcasm; and explicit false belief: the unexpected transfer task) from the authorized German-language translation of the Extended Theory-of-Mind Scale (originally by Wellman & Liu 2004, Peterson et al. 2012; translated by Henning et al. 2013). For the Strange Stories, we used a further combined score of performance across pre-tested German-language translations of Happé’s Strange Stories (following Happé 1994 and White et al. 2009; pre-tested German translations by Rakoczy et al. 2012, 2018; Ebert 2020). The eight Strange Stories were evenly divided between and corresponded to four themes: double bluff, white lie, persuasion (deception), and misunderstanding. A neutral (*Kannst du es mir noch ein bisschen genauer sagen?* [Can you say that a bit more precisely for me?]) and a mental (e.g., *Was glaubt Simon, was Moritz denkt?* [What does Simon believe Moritz thinks?]) follow-up question was prepared for each story. These were used by the experimenter when it was necessary to seek further clarification of the children’s responses. Audio files for each of the Strange Stories had been pre-recorded by a native speaker and were played twice for the children. While they listened to and were asked about the Strange Stories, they were shown a grey neutral screen with a small black cross in the center. This was to hold their focus and prevent them being distracted by anything else on the screen. Minor adaptations to the stimuli for both the ToM-scale and Strange Stories were made to fit them to the digital or modern context; full details of these adaptations are in the appendices. A second coder independently scored 28.6% (a randomly selected 10 of the 35 total participants) of the children’s answers and interrater reliability was calculated (ToM-scale: Cohen’s $\kappa = 1.0$; Strange Stories: Cohen’s $\kappa = .92$).

3.4.9. Metaphor comprehension at 9;0

The children were tested at age 9;0 on their understanding of selected age-appropriate metaphors. These were embedded within short sentences to provide context (e.g., *Mein Herz ist gebrochen.* [*My heart is broken*]). This context-embedded approach is in line with existing work (cf. Babarczy et al. 2019, Özçalışkan 2005). The metaphors we used were previously utilized in a study by Vogt & Indefrey (2017) who permitted us to use them. Their work was only published and made available in the German language. Since it strongly informed our study’s measure of metaphorical understanding, we describe and summarize their work for the wider international audience in more detail within our appendices. In our study we followed Vogt & Indefrey’s (2017) procedure but adapted and shortened their original list of metaphors. We conducted our own pilot study where we surveyed 30 adult native speakers by presenting them each of the 11 categories of metaphor defined by Vogt & Indefrey (2017), for which there were 5 metaphors each. We then asked these respondents to identify which they felt were the two most common metaphors from each category. After totaling their responses, we eliminated the extremes (first and least most common) and selected the second most common and second least common metaphors from each category to form our stimuli list comprised of 22 metaphors in total. We followed this procedure to shorten the stimuli list to an appropriate length for our study but remain representative of each different category of metaphor. We considered this important given that Vogt & Indefrey (2017) found each category varied in difficulty across the age range of 9-11 years. To achieve consistency across our sample, each metaphor had been pre-recorded by a native speaker and each metaphor’s audio file was played twice for the child. The experimenter then asked the child: (1) if one could say such a sentence, (2) whether they had heard the sentence before, (3) whether they understood it, and if the child answered positively, and (4) whether they could explain the sentence in their own words. If the child answered negatively, they were asked what they thought the sentence might mean. The lead-up questions were developed by the task's creators and we decided to keep them in our study to help acclimatize the children to the task. We aimed to give them time to process the metaphors instead of immediately answering that they did not know. While the children listened to and were asked about the metaphors, they were shown the same grey neutral screen with a fixation cross used during the Strange Stories task. The children's answers were recorded, transcribed and their explanation (or idea) about what the metaphor meant was scored as either correct or incorrect (0 or 1) according to a pre-prepared list of target explanations. The other lead-up questions were not used in our analyses. A second coder scored 28.6% (a randomly selected 10 of the 35 total participants) of the children's responses independently to enable computation of interrater reliability (Cohen's κ = .79).

4. Approach to analysis

The analyses reported here were conducted using the software IBM SPSS 25 for Windows. They focus on establishing whether early gesture development (index-finger pointing at age 1;0) can be identified as a predictor of outcomes in language, ToM, and metaphor comprehension skills at age 9;0, and whether it is mediated by iconic gesture comprehension or language skills at age 3;0. To address this focus, we conducted a sequence of regression and mediation analyses. The regression analyses were stepwise with the independent variables: index-finger pointing at 1;0, socioeconomic status (SES) of the family at 1;0, and nonverbal IQ measured at 3;6. Dependent variables were ToM-scale, ToM Strange Stories, metaphor comprehension, productive vocabulary, productive grammar, and receptive grammar at 9;0. Predictor variables were included in the final models if they could significantly improve the models' abilities to predict the outcome variables. Mediation analyses were conducted utilizing the PROCESS macro developed by Hayes (2018). They incorporated index-finger pointing at 1;0 as the predictor variable and iconic gesture comprehension at 3;0 or language at 3;0 as mediator variables. The outcome variables at 9;0 were productive vocabulary, productive grammar, receptive grammar, the ToM-scale, the ToM Strange Stories, and metaphor comprehension. No further variables were taken into account. The PROCESS macro employs ordinary least squares regression to generate unstandardized path coefficients for total, direct, and indirect effects. To compute confidence intervals and inferential statistics, a bootstrapping technique with 5,000 samples, along with heteroscedasticity-consistent standard errors (Davidson & MacKinnon 1993), was employed. Significance of effects was determined based on whether the confidence interval did not include zero.

Table 2. Descriptive statistics for mean performance, standard deviation and score range across all tests with sample size indicated in parentheses after each test.

Test (Sample Size)	*M, SD* (Score Range)
Index-finger pointing (decorated room) at 1;0 (yes or no; *N* = 35)	Yes *n* = 27 No *n* = 8
Communicative motives for pointing at 1;0 (*N* = 35)	
Mean number of declarative-expressive points	1.04, 1.17 (0.00–3.75)
Mean number of declarative-informative points	0.23, 0.33 (0.00–1.00)
Mean number of imperative points	2.16, 1.63 (0.25–6.50)
Iconic gesture comprehension at 3;0 (maximum of 16; *N* = 33)	7.39, 2.60 (1–13)
Composite value of language skills at 3;0 (factor loading; *N* = 35)	0.06, 1.02 (-1.94–2.07)
Nonverbal IQ at 3;6 (standard IQ-score; *N* = 32)	106.94, 13.14 (80–129)
Productive vocabulary at 9;0 (standard T-score; *N* = 35)	53.03, 7.76 (38–68)
Productive grammar at 9;0 (standard T-score; *N* = 35)	59.29, 7.54 (44–74)
Receptive grammar at 9;0 (standard T-score; *N* = 35)	53.71, 8.47 (37–69)
ToM (Theory-of-Mind scale: hidden emotion, sarcasm, explicit false belief) at 9;0 (maximum of 3; *N* = 34)	2.09, 0.67 (1–3)
ToM (Strange Stories) at 9;0 (maximum of 16; *N* = 33)	12.09, 2.54 (4–16)
Metaphor comprehension (explanation of sentences containing metaphors; maximum of 22; *N* = 35)	9.37, 3.53 (1–19)

Table 3. Pearson's correlations between all considered variables.

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.
1. Socioeconomic Status at 1;0	–	.117 (34)	-.175(34)	-.100 (34)	-.121(34)	.548** (32)	.435* (34)	.376 (31)	.428* (34)	.322 (34)	.532** (34)	.265 (33)	.473* (32)	.372 (34)
2. Index-finger pointing at 1;0	.117 (34)	–	.282 (35)	.228 (35)	.414* (35)	.557** (33)	.647** (35)	.176 (32)	.483* (35)	.305 (35)	.299 (35)	.180 (34)	.285 (33)	.586** (35)
3. Declarative-expressive pointing at 1;0	-.175(34)	.282 (35)	–	.316 (35)	.653** (35)	.289 (33)	.196 (35)	.289 (32)	-.099 (35)	-.205 (35)	.219 (35)	.195 (34)	-.113 (33)	.278(35)
4. Declarative-informative pointing at 1;0	-.100 (34)	.228 (35)	.316 (35)	–	.198 (35)	.312 (33)	.102 (35)	-.040 (32)	-.001 (35)	-.074 (35)	-.024 (35)	.249 (34)	-.061 (33)	.113(35)
5. Imperative pointing at 1;0	-.121(34)	.414*(35)	.653** (35)	.198 (35)	–	.247 (33)	.379 (35)	.144 (32)	.091 (35)	-.091 (35)	.316 (35)	.055 (34)	.082 (33)	.296 (35)
6. Iconic gesture comprehension at 3;0	.548** (32)	.557** (33)	.289 (33)	.312 (33)	.247 (33)	–	.637** (33)	.257 (31)	.649** (33)	.417* (33)	.488* (33)	.445* (32)	.345 (31)	.770** (33)
7. Composite language skills at 3;0	.435* (34)	.647** (35)	.196 (35)	.102 (35)	.379 (35)	.637** (33)	–	.349 (32)	.686** (35)	.411* (35)	.596** (35)	.402* (34)	.623** (33)	.689** (35)
8. Nonverbal IQ at 3;6	.376 (31)	.176 (32)	.289 (32)	-.040 (32)	.144 (32)	.257 (31)	.349 (32)	–	.196 (32)	.298 (32)	.416* (32)	-.043 (31)	.188 (30)	.236 (32)
9. Vocabulary at 9;0	.428* (34)	.483* (35)	-.099 (35)	-.001 (35)	.091 (35)	.649** (33)	.686** (35)	.196 (32)	–	.575** (35)	.545** (35)	.446* (34)	.542** (33)	.641** (35)
10. Productive Grammar at 9;0	.322 (34)	.305 (35)	-.205 (35)	-.074 (35)	-.091 (35)	.417* (33)	.411* (35)	.298 (32)	.575** (35)	–	.437* (35)	-.085 (34)	.383 (33)	.265 (35)
11. Receptive Grammar at 9;0	.532** (34)	.299 (35)	.219 (35)	-.024 (35)	.316 (35)	.488* (33)	.596** (35)	.416* (32)	.545** (35)	.437* (35)	–	.144 (34)	.492* (33)	.418* (35)
12. ToM-Scale at 9;0	.265 (33)	.180 (34)	.195 (34)	.249 (34)	.055 (34)	.445* (32)	.402* (34)	-.043 (31)	.446* (34)	-.085 (34)	.144 (34)	–	.271 (32)	.492* (34)
13. Strange Stories at 9;0	.473* (32)	.285 (33)	-.113 (33)	-.061 (33)	.082 (33)	.345 (31)	.623** (33)	.188 (30)	.542** (33)	.383 (33)	.492* (33)	.271 (32)	–	.502* (33)
14. Metaphor Comprehension at 9;0	.372 (34)	.586** (35)	.278(35)	.113(35)	.296 (35)	.770** (33)	.689** (35)	.236 (32)	.641** (35)	.265 (35)	.418* (35)	.492* (34)	.502* (33)	–

Abbreviations: ToM, theory of mind.

Socioeconomic status at 1;0, index-finger pointing at 1;0, declarative-expressive pointing at 1;0, declarative-informative pointing at 1;0, imperative pointing at 1;0, iconic gesture comprehension at 3;0, composite language skills at 3;0, nonverbal IQ at 3;6, productive vocabulary at 9;0, productive grammar at 9;0, receptive grammar at 9;0, ToM-scale at 9;0, Strange Stories at 9;0, metaphor comprehension at 9;0 (Pearson's r, with sample size indicated in parentheses after each coefficient; all *p* values have been corrected for multiple comparisons using the Bonferroni-type Benjamini & Hochberg [1995] method).

Prior to addressing our analyses, we present as first results the descriptive statistics of each test used across the analyses (Table 2) and the results of the Pearson's correlational analyses identifying the correlations and intercorrelations between all of the predictor and outcome variables we address in this study (Table 3). We followed existing argumentation regarding appropriate correction methods when carrying out multiple comparisons (Armstrong 2014, Benjamini & Hochberg 1995, Benjamini & Yekutieli 2001, Bland & Altman 1995, Jafari & Ansari-Pour 2019). We therefore employed the Bonferroni-type Benjamini & Hochberg (1995) method for controlling the false discovery rate and correcting our *p* values across the full family of 91 tests (Table 3 presents these corrected *p* values). Due to the longitudinal nature of the study and the number of tasks conducted with each participant, a small minority of cases had missing data. Except for the calculation for the composite language measure at 3;0 years, pairwise deletion was used across all the analyses. Sample size is therefore reported for each analysis.

5. Results

The results of the descriptive statistics (Table 2) indicate that the children as a group demonstrated average performances in the areas of language and cognition. The Pearson's correlational analyses (Table 3) were corrected for multiple comparisons using the Bonferroni-type Benjamini & Hochberg (1995) method.[1] Index-finger pointing at 1;0 was correlated with productive vocabulary skills at age 9;0 and metaphor comprehension at age 9;0, but not with performance on the ToM-scale, the ToM Strange Stories or grammatical skills at age 9;0. The different communicative motives for pointing at age 1;0 were not correlated with any outcome variables at age 9;0 and will thus be excluded from further analysis. Iconic gesture comprehension at age 3;0 was correlated with all language skills, performance on the ToM-scale and metaphor comprehension at age 9;0 but not performance on the ToM Strange Stories task. The composite value of language skills at 3;0 (factor loading; the calculation of this value is explained in the Methods section) was correlated with all outcome variables at age 9;0. Performance on both measures of ToM correlated with metaphor comprehension at age 9;0.

5.1. Do gestural skills at 1;0 predict ToM, metaphor comprehension, and language at 9;0?

We begin chronologically with the earliest point of gestural competence, and consider whether gestural skills at age 1;0 could predict ToM, metaphor comprehension and language skills at age 9;0. To do so, we conducted six separate stepwise regression analyses (Figure 2) with index-finger pointing at age 1;0 (yes or no) as the predictor variable for all outcome variables at age 9;0 (ToM-scale, ToM Strange Stories, metaphor comprehension, productive vocabulary, productive grammar, and receptive grammar). We further included SES at age 1;0 and nonverbal IQ at age 3;6 as potential control variables in the analyses, as each of these variables had also correlated with some of the outcome variables at age 9;0 (shown previously in Table 2). The ability to point with the index finger at age 1;0 explained 41% of the variance in metaphor comprehension skills at age 9;0 (Figure 2). In a joint model with SES at age 1;0, it also explained 41% of the variance in productive vocabulary skills at 9;0. ToM skills were not predicted by index-finger pointing, and the nonverbal IQ of the children, measured at 3;6 years was not predictive of any outcome variable.

After applying the Bonferroni-type Benjamini & Hochberg (1995) correction method to control for the false discovery rate across multiple testing, only the following regression models from Figure 2 remained significant at the alpha level <.05: (1) index-finger pointing at 1;0 predicted metaphor comprehension at 9;0, and (2) SES at 1;0 predicted receptive grammar at 9;0. However, the multiple

[1] All variables were checked for the existence of outliers (+/-2 SD). We only found three outliers, one each for the variables declarative-expressive points, ToM Strange Stories, and metaphor comprehension. Since we were particularly interested in individual differences and below-average performances, we did not exclude these children and their data from the analyses. We further checked for any influence by repeating the analyses excluding this data; these produced comparable results.

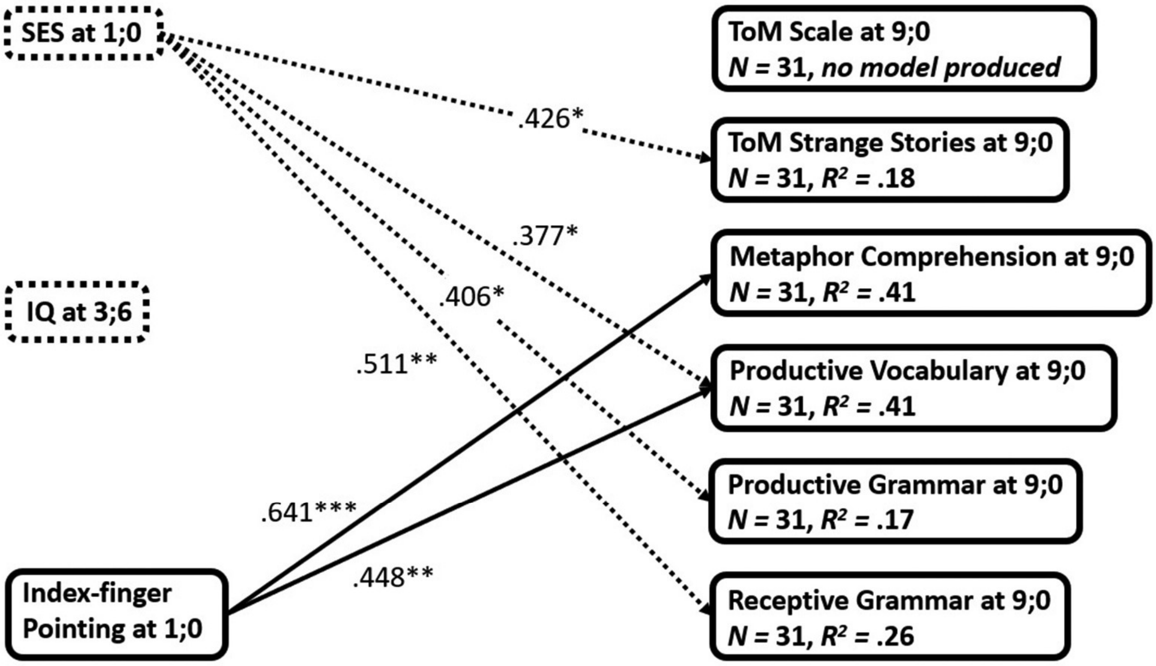


Figure 2. Final models of 6 stepwise regression analyses with the independent variables index-finger pointing at 1;0, socioeconomic status (SES) of the family at 1;0, and nonverbal IQ measured at 3;6. Dependent variables were theory of mind (ToM)-scale, ToM Strange Stories, metaphor comprehension, productive vocabulary, productive grammar, and receptive grammar at 9;0. Predictor variables were included in the model if they could significantly improve the models' abilities to predict the outcome variables. Presented are standardized betas of the predictor variables and the R^2 of the final models, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. The depicted models are pre-correction for multiple testing; post-correction using the Bonferroni-type Benjamini & Hochberg (1995) method, only index-finger pointing at 1;0 remains significantly predictive for metaphor comprehension and productive vocabulary at 9;0, and SES at 1;0 remains significantly predictive for receptive grammar at 9;0.

regression analysis was also able to produce an alternative single predictor model for index-finger pointing at 1;0 which explained 27% of the variance in productive vocabulary outcomes at 9;0. This remained significant after the correction for multiple testing ($n = 31$, $\beta = .518$, $R^2 = .27$, $F(1,29) = 10.631$, $p = .003$). Conducting power analyses for the regression models with a coefficient of determination of $R^2 = .41$, a statistical power of .9 and a significance level of $\alpha = .05$ yielded a required sample size of $n = 22$ for a significant overall model with 2 predictors and $n = 18$ for a significant overall model with 1 predictor. IQ at 3;6 was rejected by all our stepwise regression models but a 3-predictor model would have required a sample of $n = 25$. For the corrected single-predictor regression model with the coefficient of determination of $R^2 = .27$, a sample of $n = 31$ would be required. Since our sample size was $n = 31$ for these analyses, the threshold for a significant effect is reached.

5.2. Do gestural and language skills at 3;0 mediate the relation between index-finger pointing at 1;0 and ToM, metaphor comprehension, and language skills at 9;0?

Although index-finger pointing at age 1;0 was found to be predictive of only vocabulary skills and metaphor comprehension at age 9;0, we conducted mediation analyses in line with argumentation by Zhao et al. (2010) on the lack of need for a direct effect. These analyses used index-finger pointing at 1;0 as the predictor and iconic gesture comprehension at age 3;0 as the mediator (Figure 3) for all outcome variables at age 9;0 (ToM-scale, ToM Strange Stories, metaphor comprehension, productive vocabulary, productive grammar, and receptive grammar). Control variables were not included in the mediation analyses. We carried out the same analyses with language skills at age 3;0 as the mediator variable (Figure 4) instead of iconic gesture comprehension. This is because both variables have been found to mediate the relation between index-finger pointing at age 1;0 and language skills at age 6;0 (Lüke et al. 2020). Fritz & MacKinnon's (2007) simulation requires a sample size of $N = 34$ for

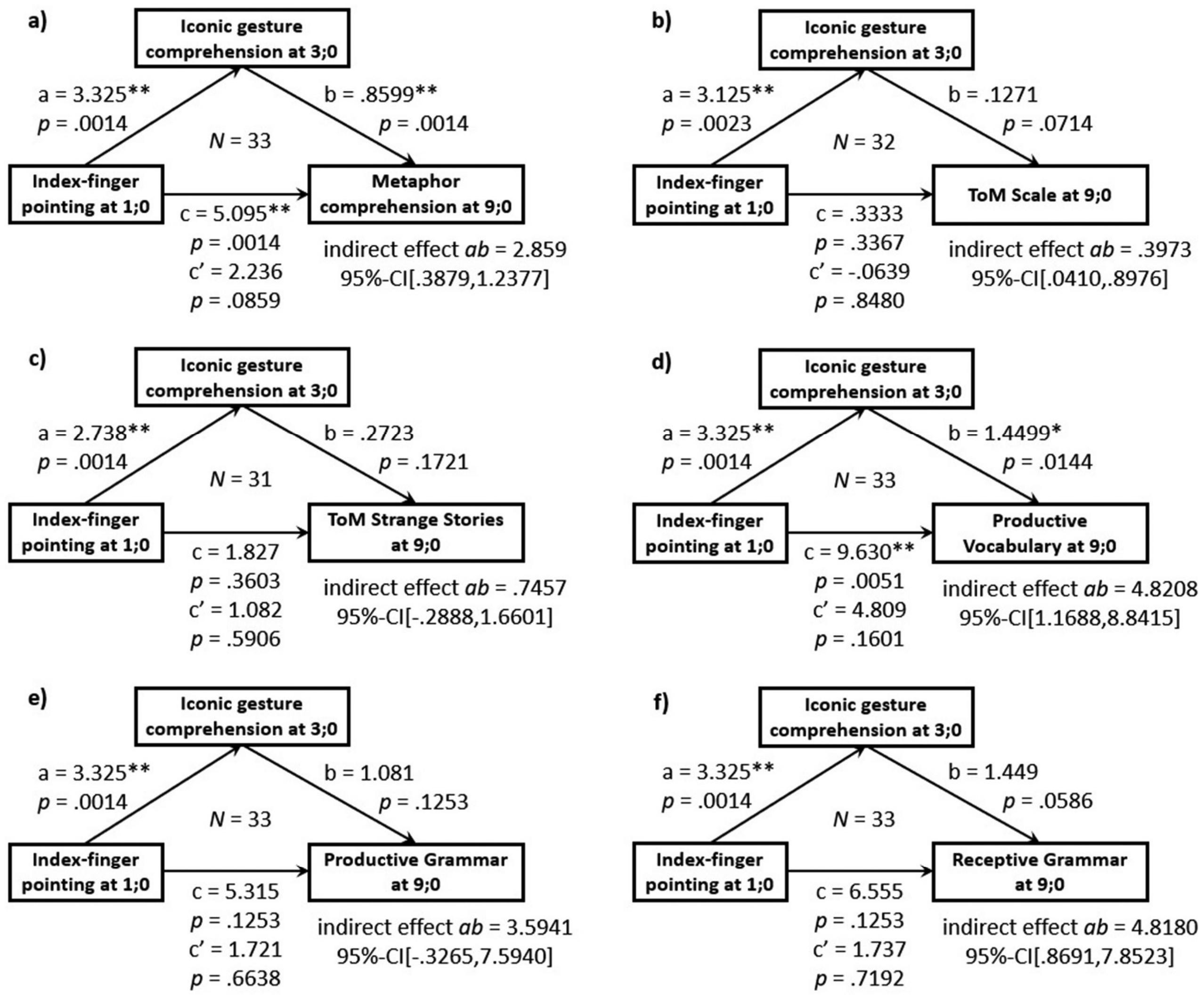


Figure 3. Models of index-finger pointing at 1;0 as a predictor of performance on all outcome variables at 9;0 (theory of mind [ToM]-scale, ToM Strange Stories, metaphor comprehension, productive vocabulary, productive grammar, and receptive grammar), mediated by iconic gesture comprehension at 3;0 (*$p < .05$, **$p < .01$, ***$p < .001$; p values shown have been adjusted using the Bonferroni-type Benjamini & Hochberg [1995] method).

mediation models with large effects for the α and β paths. We found large effect sizes for all but three analyses (Figure 3b, 3c, 4b), for which a much bigger sample size would then be appropriate. The results of the mediation analyses showed that both iconic gesture comprehension and language skills at 3;0 mediate the relation between index-finger pointing at 1;0 and metaphor comprehension (Figure 3a, 4a), performance on the ToM-scale (Figure 3b, 4b), productive vocabulary skills (Figure 3d, 4d) and receptive grammar skills at 9;0 (Figure 3f, 4f). Only language skills at 3;0 mediate the relation between index-finger pointing at 1;0 and performance on the ToM Strange Stories task at 9;0 (Figure 4c). No mediation effect of iconic gesture comprehension or language skills at 3;0 was found for productive grammar skills at 9;0.

6. Discussion

In our longitudinal study, we explored whether early gestural skills (index-finger pointing at age 1;0) could predict 9-year-olds' later performance on tasks measuring ToM (ToM-scale and Strange Stories) and metaphor comprehension as well as language competencies (productive vocabulary, productive grammar, and receptive grammar). We further evaluated whether this relation could be mediated by

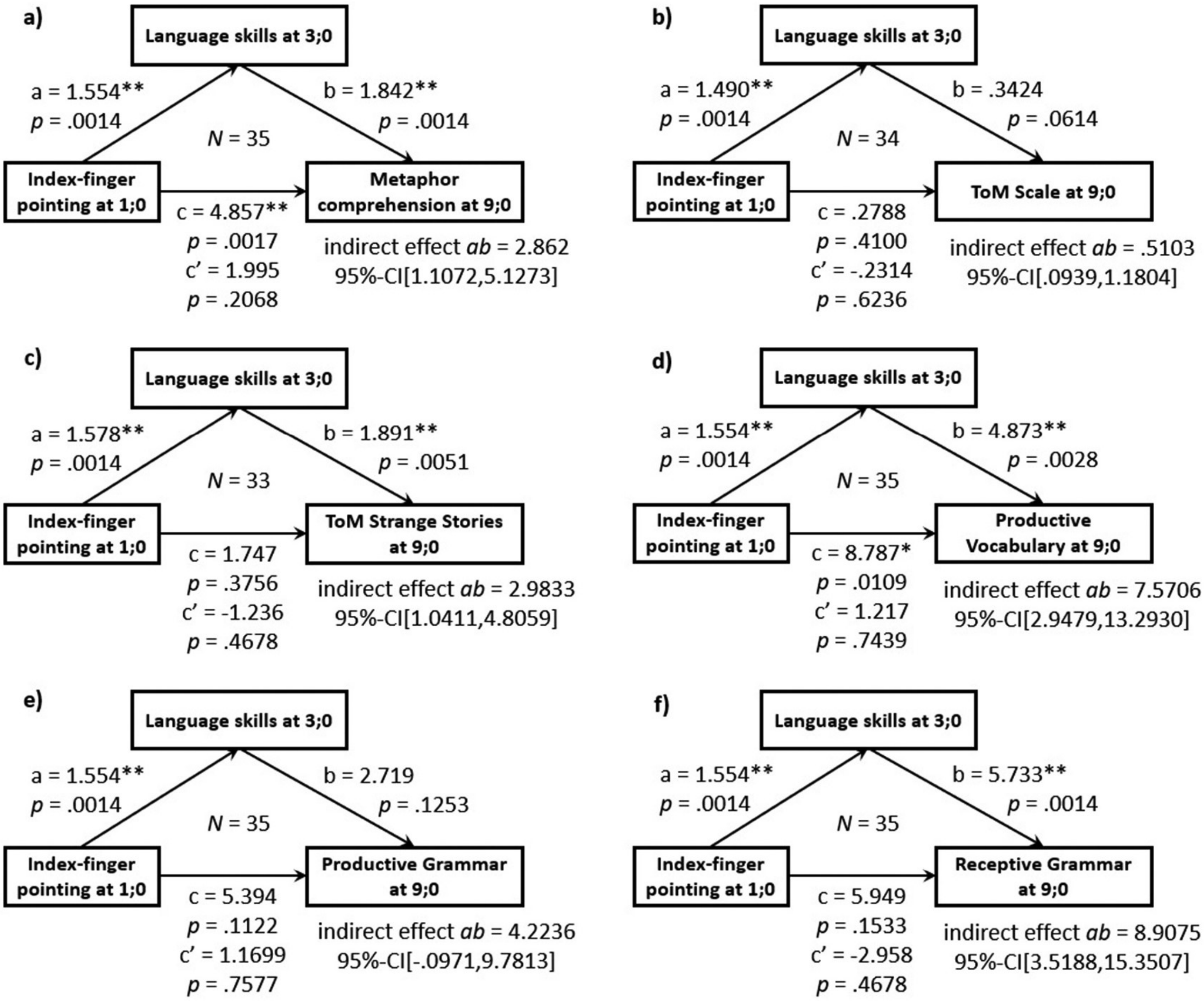


Figure 4. Models of index-finger pointing at 1;0 as a predictor of performance on all outcome variables at 9;0 (theory of mind [ToM]-scale, ToM Strange Stories, metaphor comprehension, productive vocabulary, productive grammar, and receptive grammar), mediated by language skills at 3;0 (*$p < .05$, **$p < .01$, ***$p < .001$; p values shown have been adjusted using the Bonferroni-type Benjamini & Hochberg [1995] method).

another—simple-to-measure and nonverbal—early gestural skill (iconic gesture comprehension at age 3;0) in comparison to language skills (also measured at age 3;0).

In stepwise regression models involving the potential control variables SES at 1;0 and nonverbal IQ at 3;6, we found that index-finger pointing at 1;0 was directly predictive of metaphor comprehension and vocabulary skills at age 9;0. This strong association between index-finger pointing and metaphor comprehension as well as productive vocabulary skills was also reflected in the statistically significant correlational analyses.

Although we did not find similar direct effects in relation to the other outcome variables at 9;0, we did find indirect connections between this early marker of gestural skills and later, more complex skills. We considered index-finger pointing at 1;0 within mediation models using either iconic gesture comprehension at 3;0 or language skills at 3;0 as the mediator. Results from both mediation pathways showed that index-finger pointing at 1;0 indirectly predicted metaphor comprehension, performance on one ToM task (ToM-scale), productive vocabulary skills, and receptive grammar skills at 9;0. Via only the language skills at 3;0 mediation pathway, it could further indirectly predict performance on the other ToM task (Strange Stories) at 9;0. No predictive effects were found for productive grammar skills at 9;0 via either mediation pathway. No prior work has ever identified this across this length of time or with these mediators.

Our results highlight that alongside spoken language skills, a developing understanding of gestures (index-finger pointing and iconic gestures) is also an important and valid predictor of children's ongoing development of ToM and metaphor comprehension skills. Although we did not find many direct effects of index-finger pointing at 1;0 on the outcome variables at 9;0, we did find that the mediation models showed indirectly predictive relations (either via iconic gesture comprehension at 3;0 or composite language skills at 3;0) for almost all of the outcomes at 9;0. This strong albeit indirect connection between whether or not children were able to point with the index-finger at 1;0 (not the communicative motive for pointing) and outcomes at 9;0 in ToM, metaphor comprehension, and language skills suggests that there is an important relation between these domains, which few have previously considered. This is particularly striking when considering the mediation models where another gestural competency—iconic gesture comprehension at 3;0—functioned as the mediator. These findings indicate that a nonverbal test of iconic gesture comprehension at 3;0 could potentially be a quick and simple alternative to lengthy standardized language tests for predicting later language, ToM and metaphor comprehension skills. This should however be verified within a larger sample.

These results support the theory that gestures and oral language form an integrated communication system (e.g., Kita & Özyürek 2003). They would also fit with the idea that using and processing gestures—as a form of 'embodied' language, embedded within a communicative context—may draw on both embodied and embedded cognition. Some theories have situated early ToM in embodied cognitive processes (Fenici 2012) and, in line with this, experimental work has indicated that more general pointing (e.g., initiating joint attention) is predictive of early ToM development (Brandone & Stout 2023, Cochet et al. 2017, Colonnesi et al. 2008, Sodian & Kristen-Antonow 2015). Fenici (2012) also proposed that later ToM processes are scaffolded and supported by the external structure of language, which we believe could additionally integrate gesture. The findings of our mediation models could support both the embodied view of early ToM and embedded view of later ToM. Our predictor variable index-finger pointing at 1;0 could reflect early embodied ToM abilities when initiating joint attention. Our mediator variables (language skills at 3;0 or iconic gesture comprehension at 3;0) might then bridge the gap between embodied and embedded ToM processes, since our outcome variables at 9;0 of the ToM-scale and Strange Stories are embedded in language. Work identifying that iconic gestures support FB reasoning could further substantiate this notion (Carlson et al. 2005), and these early intuitions might be worth future exploration.

The combination of embodiment and embeddedness might also be paralleled within metaphor comprehension. This domain appears to straddle the areas of language and ToM, since specific metaphors draw differentially on these skills depending upon their content (Pronina et al. 2023, Tonini et al. 2023). Recent work proposing metaphorical embodiment (Khatin-Zadeh, Farsani, et al. 2023) could support this idea. Conducting a review of prior studies, Khatin-Zadeh and colleagues concluded that the degree of involvement of sensorimotor systems when mentally simulating metaphoric actions differs across metaphors, depending upon their base's sensorimotor strength (Khatin-Zadeh, Farsani, et al. 2023). This resonates with work identifying that the link between ToM and metaphor comprehension is stronger when metaphors involve mental rather than physical aspects (Lecce et al. 2019, Tonini et al. 2023). It could thus be the case that some metaphors are embodied while others are embedded in language. It is possible that interlocutors might be more likely to reinforce communication of embodied metaphors (as opposed to embedded ones) through gesture use, differentially utilizing perspective-taking and language skills. Gesturally depicted metaphor schema have been found to prime comprehension of their congruent metaphors, increasing and improving judgment response times regarding metaphor sensibility (Khatin-Zadeh 2023, Khatin-Zadeh, Hu, et al. 2023). These gestural depictions may also equate to embodied realizations of metaphors, activating wider neural networks representing their meaning (Khatin-Zadeh 2023). These findings directly link gesture to metaphor and complement our own results, which also establish direct connections between early (deictic and iconic) gesture use and metaphor comprehension. Taken together, they bolster the idea of a wider general communication framework incorporating the domains of gesture, language, ToM, and metaphor. This framework may also interface principles of embodiment and embeddedness, and should be investigated in further research.

Next, we would like to concentrate on more specific aspects of our findings and will begin with those regarding the construct of ToM. Most existing work on ToM development has focused on early childhood (up to age 6) but recent work has demonstrated and argued that its acquisition is an ongoing process shaped throughout middle childhood and even into adolescence (Devine & Lecce 2021, Osterhaus & Koerber 2021, Osterhaus et al. 2022, Tonini et al. 2023). For this reason, we chose to investigate 9-year-old children's ToM abilities alongside their development of metaphor comprehension in our longitudinal study. We did this because at age 9 they fall right in the center of an influential period of ToM and more complex linguistic development during middle childhood. By considering this stage, we hoped to contribute toward closing a research gap.

In relation to ToM, we would first like to consider the role of language and gesture. In line with previous work (Astington & Baird 2005, Doherty 2008, Durrleman 2020, Durrleman & Delage 2020, Milligan et al. 2007, Osterhaus & Koerber 2021), we found that earlier language skills were relevant for later ToM performance. Unlike Colonnesi et al. (2008), Sodian & Kristen-Antonow (2015), and Cochet et al. (2017), we did not find a directly predictive role of pointing in ToM development. That being said, we did find that iconic gesture as a mediator indirectly linked pointing with ToM, which is potentially analogous to Carlson et al.'s (2005) work. However, we conducted two separate measures of ToM skills: the ToM-scale and the Strange Stories task; and these two measures manifested differentially within our results. Iconic gesture comprehension at 3;0 mediated the relation between index-finger pointing at 1;0 and performance on the ToM-scale at 9;0, which included an explicit FB task; this task type was also used by Carlson et al. (2005). The composite measure of children's language skills at 3;0, however, mediated the relation between index-finger pointing at 1;0 and performance on both the ToM-Scale and the Strange Stories at 9;0. We also observed that of the language skills, only vocabulary at age 9;0 was correlated with performance on the ToM-scale at age 9;0. While it was also the strongest correlate with performance on the Strange Stories task at 9;0, receptive grammar at 9;0 was also a relevant factor. This suggests that the tasks entailed different linguistic demands.

Such findings are interesting given the contrasts in prior work (Astington & Baird 2005, De Villiers & Pyers 2002, Durrleman 2020, Durrleman & Delage 2020, Happé 1995a, Milligan et al. 2007, Osterhaus & Koerber 2021). They indicate that different measures of ToM should perhaps be re-evaluated and considered in connection with different linguistic competencies. For example, the longer and more complex texts involved in the Strange Stories task might require the participant to have a stronger grasp of grammatical intricacies. In contrast to the purely auditory Strange Stories task, the visual supports used in the ToM-scale tasks might have lightened the load of linguistic processing required. They may even have tapped visual processing skills also required for the processing of iconic gesture. Moreover, our sample's performances on the ToM-scale subtests and the Strange Stories did not actually correlate with each other. Considering our finding that iconic gesture comprehension and general language skills played different roles in mediating these measures' outcomes, this raises questions about whether these measures test the same or even similar aspects of ToM. We would like to emphasize here that future work should take this into account when investigating relations between children's ToM and other skills. Different measures of ToM test different aspects of mentalization and perspective-taking, which may in turn strongly influence one's results.

The next point focuses on the specific contribution of iconic gesture to ToM development. As highlighted by Carlson et al. (2005), iconic gesture possesses several features which are also relevant to ToM development. These include its ability to convey information not present in accompanying speech and its provision of alternative routes for representing modes of thought. A key point is that such gestures are not arbitrary symbols; they present embodied, visuospatial formats of their verbal representation counterparts, which may be easier for children to access. Just like spoken language, iconic gestures can also represent some other entity, object or abstract concept. Prior work has already linked children's developing understanding of the nature of pictorial, linguistic and metalinguistic representations to their acquisition of ToM (Doherty 2000, Doherty 2008, Doherty and Perner 1998). Iconic gesture comprehension has also been separately linked to developing language skills (cf. Lüke et al. 2020) while these have long been implicated in ToM development (cf. Milligan et al. 2007). Bringing everything together, it follows logically that iconic gesture would also contribute to ToM development. More targeted work with larger populations is needed, however, to go beyond our study's scope and pinpoint exactly how it does so.

We will now turn to metaphor comprehension and seek to address the specific roles played by index-finger pointing, iconic gesture comprehension, language skills, and ToM in its development. Our results indicated that metaphor comprehension at 9;0 was strongly correlated with index-finger pointing at 1;0, iconic gesture comprehension at 3;0, and language skills at 3;0. However, it also correlated with a number of other variables under consideration (vocabulary at 9;0, receptive grammar at 9;0; ToM-scale at 9;0, and Strange Stories at 9;0). This implies that metaphor comprehension draws on many different skills and experiences. This fits with the existing work identifying several different processes involved in metaphor comprehension (Deckert et al. 2019, Kalandadze et al. 2018, Khatin-Zadeh, Farsani, et al. 2023, Pronina et al. 2023, Tonini et al. 2023, Vulchanova et al. 2019).

Iconic gesture comprehension at 3;0 and language skills at 3;0 also both mediated a relation between index-finger pointing at 1;0 and metaphor comprehension at 9;0. There are a number of possible factors which could be contributing to this finding: (1) Gesture emerges earlier than language and since we used measures of the children's abilities at ages 1;0 and 3;0, their gesture skills at this point might have been slightly more sophisticated than their language. This may have led to a potentially more precise measurement. (2) Gesture is also a representational form of communication and may function as a type of embodied, visual metaphor for the spoken language it accompanies. (3) Metaphor and gesture comprehension might both tap similar but wider visual and perceptual abilities than those required by more general language. Larger and more highly targeted future studies are needed to attempt to weigh up these potentially contributory factors and more finely evaluate the different connections.

Our results also showed that specifically index-finger pointing at 1;0 could explain a large amount of the variance in metaphor comprehension at age 9;0, measured 8 years later. Index-finger pointing at 1;0 also predicts the mediator variable iconic gesture comprehension at 3;0 (Lüke et al. 2020), so first we turn to the question: What is the specific role here of the index-finger? The significant correlation between index-finger pointing at 1;0 and metaphor comprehension at 9;0 in our sample points to a cognitive explanation. Index-finger pointing is itself an indicator of cognitive development and a fundamental turning point in infants' understanding of communicative intentions (Liszkowski & Tomasello 2011). Other work (Lüke, Grimminger, et al. 2017) has also indicated that the hand shape of pointing (index finger vs. whole hand) appears to be particularly important for the development of general linguistic competencies, rather than the motives behind it. The shape of an index-finger point is far less ambiguous than a whole-hand point. It signifies the speaker's intention to single something out for the interlocutor much more clearly; in this way it is not merely deictic but also functioning in a somewhat representational capacity. Children's growing social cognitive abilities and understanding of communicative representations might be scaffolded by their developing awareness that (1) an index-finger point stands for something distinctive communicatively, (2) variants of gestures can be perceived by their interlocutor to differential effect, and (3) gestures must consequently be designed for their interlocutor's benefit. Iconic gestures are much further along the representational spectrum than index-finger pointing and have been recognized as metaphoric in themselves (Khatin-Zadeh 2023; McNeill 1992, 2014). It is therefore unsurprising that we find such a strong link between iconic gesture and metaphor comprehension. Both require representational "as-if" processing and the recognition that something can stand for something else, tap understanding of communicative intentions, and can become conventionalized within speech communities (cf. Holyoak & Stamenković 2018, Kendon 2004, McNeill 1992, Schnell 2007).

Next, we move onto the role of specific linguistic competencies in developing metaphor comprehension skills. We are particularly interested in why metaphor comprehension at age 9;0 in our sample correlated strongly with vocabulary at the same age but only weakly with receptive grammar and not at all with productive grammar. Previous work has found that language skills are associated with metaphor comprehension (Kalandadze et al. 2018) but also more specifically that semantic knowledge or skills are key to this process (Deckert et al. 2019, Vulchanova et al. 2019). It is likely that this connection between productive vocabulary and metaphor comprehension within our sample is due to our method of testing the children's vocabulary (P-ITPA; Esser et al. 2010). Eliciting vocabulary with this method requires children to think about semantics to call up the correct word ("I'm thinking about something with feathers, what could that be?"). This stands in contrast to purely picture naming tasks, which are lexical. Further research is needed to establish how exactly vocabulary skills relate to metaphor comprehension in comparison with other linguistic competencies and to clearly establish any potential effects of measurement method.

Looking at our results, the exact contribution of ToM to metaphor comprehension remains an open question. Like Lecce et al. (2019) and Tonini et al. (2023), we did find correlations between metaphor comprehension and ToM performance in the tasks we used with 9-year-olds but these effects were much stronger between metaphor comprehension and gestural or linguistic competencies. This may be a consequence of the types of metaphors we used in our stimuli. Lecce, Tonini, and colleagues found that the link between metaphor and ToM was strongest when metaphors involved mental aspects, and, by design, our metaphors were drawn from a wide variety of categories. Another important possibility has been identified in recent work: Specific pragmatic skills may fall differentially along a spectrum between ToM and structural language ability interfaces (Pronina et al. 2023). Nonliteral, social pragmatic skills (e.g., sarcasm, metaphor comprehension) requiring additional consideration of a speaker's intended meaning and mental state have been linked to ToM ability (Pronina et al. 2023). Linguistic pragmatic skills (e.g., socially appropriate speech acts) appear to draw on structural language skills instead (for a discussion, see Pronina et al. 2023). Tonini et al. (2023) suggested that ToM might differentially support metaphor comprehension in different tasks. This may transpire by promoting access to mental state terms or by improving processing of context, serving as a form of "springboard" in middle childhood to acquire more sophisticated pragmatic skills (Tonini et al. 2023, p. 11). These three studies' outcomes (Lecce et al. 2019, Pronina et al. 2023, Tonini et al. 2023) could justify our finding that metaphor comprehension was linked slightly more closely to structural language than ToM skills. More in-depth research is needed to pinpoint how each gestural and linguistic competency discussed above contributes differentially to ToM and metaphor comprehension skills as well as their relation to one another.

A final issue worth bearing in mind is the impact of participant background and the role of SES. Prior work (e.g., McGillion et al. 2017) has shown that SES can predict the course of children's language development. In our sample, SES at age 1;0 correlated with performance on the ToM Strange Stories task at 9;0 as well as the productive vocabulary and receptive grammar measures at age 9;0. This suggests that the impact of SES on early linguistic (and gestural) development might subsequently influence the development of later language-related competencies such as ToM and metaphor comprehension. This question could be considered in future work.

7. Limitations

This study's strength lies in its longitudinal perspective and detailed consideration of the participating children's development, evaluated across many years and points of data collection, using a wide variety of measures. Nevertheless, there are also some limitations that we would like to address. First, due to the small sample size ($N = 35$), we were unable to undertake cross-lagged panel analyses and can therefore only draw tentative conclusions about the developmental picture of children's linguistic, socio-cognitive, and social pragmatic development.

Second, eight of the 35 children participating in our study were also identified as late talkers at age 2;0. Of these eight, two went on to receive a later diagnosis of developmental language disorder between ages 3;0-6;0, although they no longer met criteria for diagnosis at 9;0. We decided to include these eight children within the sample reported here to represent natural variance within the population. Recruiters conducting a solely cross-sectional but similar study would not have had this historical data to draw on to consider excluding them. The children's comparative language backgrounds and profiles are however examined in further detail within another manuscript (Crawshaw et al. 2025).

Thirdly, we had originally intended prior to conducting this study that all measures would be carried out in-person. However, in response to the COVID-19 pandemic, we had to adapt our research plan and conduct partially digital and partially in-person testing. Nevertheless, the same procedure of two online sessions followed by one in-person session was consistently carried out with all participating children. This study also took place within an established longitudinal context. As the participating children and parents were already familiar with the university and many of the researchers involved in the project, the hybrid method was actually received very positively. Another mitigating advantage of the study's longitudinal context is that a developmental picture of the participating children's abilities and competencies had already been built up over many years of in-person testing. This allowed us to evaluate our findings contextually, potentially increasing the reliability of our online testing methods. A 2021 study by Escudero et al. investigated 4-year-old children's word-learning and found performance to be comparable between in-person and online conditions (also using Zoom). This recent work indicates that using online testing methods within developmental language research may be reliable. However, confirmation through further and more conclusive research with different tasks and target competencies is still needed.

8. Implications and future opportunities

The findings of this study have implications for the development of early detection methods and intervention programs. If delays in young children's index-finger pointing or iconic gesture comprehension have a long-lasting impact on their complex socio-cognitive and social pragmatic skill development even into middle childhood, then early methods could be deployed to promote acquisition of these competencies. Since ToM and social pragmatic (i.e., metaphor comprehension) skills become increasingly important for children's social and academic success as they grow older, this needs to be further addressed within the literature. Beyond focusing solely on spoken language abilities, future (longitudinal) work should also integrate measures of gestural competencies into any consideration of children's metalinguistic, social pragmatic, and socio-cognitive developmental trajectories.

Acknowledgements

We are very grateful to all the families who have continued to participate in this longitudinal study over these many years as well as to our collaborators Angela Grimminger, Katharina Rohlfing, and Ulf Liszkowski who have been involved in the project at previous stages. We are also very thankful for the work of all research assistants who assisted with or have been involved in the carrying out of our longitudinal study over all the years, the preparation of materials, or in the transcription and coding process: Jannika Böse, Annika Büssing, Olivia Espeter, Tabea Musshoff, Katharina Rademacher, Kathrin Gremplewski, Lisa Grundhöfer, Inga Hackbarth, Bianca Henze, Janina Jurjahn, Miriam Krimer, Carina Lucas, Laura Polke, Anika Queißer, Suzan Ryschka, and Annika Biewener.

Funding

The earlier stages of this longitudinal study between the years 2012-2018 were funded by the German Research Foundation (DFG) under the grant numbers [RI 898/6-1 and RI 898/6-2 to UR and LU 2116/1-2 to CL]. This latest stage beginning in 2019 was conducted independently from the previous grants with the support of UR and CL's university departments.

Data availability statement

The dataset generated for this study cannot be made publicly available as participants did not provide consent for that, however private enquiries to view an anonymized version of the dataset are welcomed and can be addressed to the corresponding author.

Disclosure statement

The authors declare and confirm that this research was conducted in the absence of any personal, commercial, or financial relationships that could be construed as a competing or potential conflict of interest.

ORCID

Camilla E. Crawshaw http://orcid.org/0000-0003-1359-8315
Carina Lüke http://orcid.org/0000-0001-8890-8686

References

Armstrong, Richard A. 2014. When to use the Bonferroni correction. *Ophthalmic and Physiological Optics* 34(5). 502–508. doi:10.1111/opo.12131

Astington, Janet Wilde & Jodie A. Baird. (Eds.). 2005. *Why language matters for theory of mind*. Oxford: Oxford University Press.

Babarczy, Anna, Andrea Balázs, & Fruzsina Krizsai. 2019. Preschoolers' metaphor comprehension. methodological issues in experimental pragmatics. *Acta Universitatis Sapientiae, Philologica*, 11(2). 133–150. https://doi.org/10.2478/ausp-2019-0017

Benjamini, Yoav & Yosef Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1). 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Benjamini, Yoav & Daniel Yekutieli. 2001. The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics* 29(4). 1165–1188. https://doi.org/10.1214/aos/1013699998

Bland, J. Martin & Douglas G. Altman. 1995. Multiple significance tests: The Bonferroni method. *BMJ* 310(6973). 170. https://doi.org/10.1136/bmj.310.6973.170

Botting, Nicola, Nicholas Riches, Marguerite Gaynor, & Gary Morgan. 2010. Gesture production and comprehension in children with specific language impairment. *British Journal of Developmental Psychology* 28(1). 51–69. https://doi.org/10.1348/026151009X482642

Brandone, Amanda C. & Wyntre Stout. 2023. The origins of theory of mind in infant social cognition: Investigating longitudinal pathways from intention understanding and joint attention to preschool theory of mind. *Journal of Cognition and Development* 24(3). 375–396. https://doi.org/10.1080/15248372.2022.2146117

Camaioni, Luigia. 1992. Mind knowledge in infancy: The emergence of intentional communication. *Early Development and Parenting* 1(1). 15–22. https://doi.org/10.1002/edp.2430010106

Camaioni, Luigia, Paola Perucchini, Francesca Bellagamba, & Cristina Colonnesi. 2004. The role of declarative pointing in developing a theory of mind. *Infancy* 5(3). 291–308. https://doi.org/10.1207/s15327078in0503_3

Carriedo, Nuria, Antonio Corral, Pedro R. Montoro, Laura Herrero, Patricia Ballestrino, & Iraia Sebastián. 2016. The development of metaphor comprehension and its relationship with relational verbal reasoning and executive function. *PLOS One* 11(3). e0150289. https://doi.org/10.1371/journal.pone.0150289

Carlson, Stephanie M., Antoinette Wong, Margaret Lemke, & Caron Cosser. 2005. Gesture as a window on children's beginning understanding of false belief. *Child Development* 76(1). 73–86. https://doi.org/10.1111/j.1467-8624.2005.00830.x

Chahboun, Sobh, Valentin Vulchanov, David Saldaña, Hendrik Eshuis, & Mila Vulchanova. 2016. Can you play with fire and not hurt yourself? A comparative study in figurative language comprehension between individuals with and without autism spectrum disorder. *PLoS One* 11(12). e0168571. https://doi.org/10.1371/journal.pone.0168571

Cochet, Hélène, Marianne Jover, Cécile Rizzo, & Jacques Vauclair. 2017. Relationships between declarative pointing and theory of mind abilities in 3-to 4-year-olds, *European Journal of Developmental Psychology* 14(3). 324–336. https://doi.org/10.1080/17405629.2016.1205975

Colonnesi, Cristina, Carolien Rieffe, Willem Koops, & Paola Perucchini. 2008. Precursors of a theory of mind: A longitudinal study. *British Journal of Developmental Psychology* 26(4). 561–577. https://doi.org/10.1348/026151008X285660

Colonnesi, Cristina, Geert Jan J. M. Stams, Irene Koster, & Marc J. Noom. 2010. The relation between pointing and language development: A meta-analysis. *Developmental Review* 30(4). 352–366. https://doi.org/10.1016/j.dr.2010.10.001

Crawshaw, Camilla E., Carina Lüke, & Ute Ritterfeld. 2025. Late talkers' language, metaphor, theory of mind, and reading skills at 9 years. *Journal of Speech, Language, and Hearing Research* 1–18. https://doi.org/10.1044/2024_JSLHR-23-00703

Davidson, Russell & James G. MacKinnon. 1993. *Estimation and Inference in Econometrics*. Oxford: Oxford University Press.

Deckert, Matthias, Michaela Schmoeger, Ines Schaunig-Busch, & Ulrike Willinger. 2019. Metaphor processing in middle childhood and at the transition to early adolescence: The role of chronological age, mental age, and verbal intelligence. *Journal of Child Language* 46(2). 334–367. https://doi.org/10.1017/S0305000918000491

Del Sette, Paola, Valentina Bambini, Luca Bischetti, & Serena Lecce. 2020. Longitudinal associations between theory of mind and metaphor understanding during middle childhood. *Cognitive Development* 56. 100958. https://doi.org/10.1016/j.cogdev.2020.100958

De Villiers, Jill G. 2007. The interface of language and theory of mind. *Lingua* 117(11). 1858–1878. https://doi.org/10.1016/j.lingua.2006.11.006

De Villiers, Jill G. & Jennie E. Pyers. 2002. Complements to cognition: A longitudinal study of the relationship between complex syntax and false-belief-understanding. *Cognitive Development* 17(1). 1037–1060. https://doi.org/10.1016/S0885-2014(02)00073-4

Devine, Rory T. & Claire Hughes. 2016. Measuring theory of mind across middle childhood: Reliability and validity of the silent films and strange stories tasks. *Journal of Experimental Child Psychology* 149. 23–40. https://doi.org/10.1016/j.jecp.2015.07.011

Devine, Rory T. & Serena Lecce. (Eds.). 2021. *Theory of mind in middle childhood and adolescence: Integrating multiple perspectives*. London: Routledge.

Di Paola, Simona, Filippo Domaneschi, & Nausicaa Pouscoulous. 2020. Metaphorical developing minds: The role of multiple factors in the development of metaphor comprehension. *Journal of Pragmatics* 156. 235–251. https://doi.org/10.1016/j.pragma.2019.08.008

Doherty, Martin J. 2000. Children's understanding of homonymy: Metalinguistic awareness and false belief. *Journal of Child Language* 27(2). 367–392. https://doi.org/10.1017/S0305000900004153

Doherty, Martin J. 2008. *Theory of mind: How children understand others' thoughts and feelings*. New York: Psychology Press.

Doherty, Martin J. & Josef Perner. 1998. Metalinguistic awareness and theory of mind: Just two words for the same thing? *Cognitive Development* 13(3). 279–305. https://doi.org/10.1016/S0885-2014(98)90012-0

Doherty, Martin J. & Josef Perner. 2020. Mental files: Developmental integration of dual naming and theory of mind. *Developmental Review* 56. 100909. https://doi.org/10.1016/j.dr.2020.100909

Durrleman, Stephanie. 2020. Mentalizing: What's language got to do with it? *Language Acquisition* 27. 255–275. https://doi.org/10.1080/10489223.2020.1769624

Durrleman, Stephanie & Hélène Delage. 2020. Training complements for belief reasoning in developmental language disorder. *Journal of Speech, Language, and Hearing Research* 1–17. https://doi.org/10.1044/2020_JSLHR-19-00075

Ebert, Susanne. 2020. Theory of mind, language, and reading: Developmental relations from early childhood to early adolescence. *Journal of Experimental Child Psychology* 191. 104739. https://doi.org/10.1016/j.jecp.2019.104739

Escudero, Paola, Gloria Pino Escobar, Charlotte G. Casey, & Kristyn Sommer. 2021. Four-year-old's online versus face-to-face word learning via ebooks. *Frontiers in Psychology* 12. 450. https://doi.org/10.3389/fpsyg.2021.610975

Esser, Günter, Anne Wyschkon, Katja Ballaschk, & Sylvana Hänsch. 2010. Potsdam-Illinois Test für Psycholinguistische Fähigkeiten – *P-ITPA [Potsdam-Illinois Test for Psycholinguistic Abilities – P-ITPA]*. Göttingen: Hogrefe, Testzentrale.

Fenici, Marco. 2012. Embodied social cognition and embedded theory of mind. *Biolinguistics*, 6(3–4). 276–307. https://doi.org/10.5964/bioling.8921

Fox, Annette V. 2013. TROG-D. Test zur Überprüfung des Grammatikverständnisses [*Test for Assessing Grammatical Understanding*], 6th edn. Idstein: Schulz-Kirchner.

Fritz, Matthew S. & David P. Mackinnon. 2007. Required sample size to detect the mediated effect. *Psychological Science* 18(3). 233–239. https://doi.org/10.1111/j.1467-9280.2007.01882.x

Goldin-Meadow, Susan. 2020. Using gesture to identify and address early concerns about language and pragmatics. *Pediatrics* 146(Supplement 3). S278–S283. https://doi.org/10.1542/peds.2020-0242G

Goldin-Meadow, Susan. 1998. The development of gesture and speech as an integrated system. *New Directions for Child Development* 79. 29–42. https://doi.org/10.1002/cd.23219987903

Graziano, Maria & Marianne Gullberg. 2018. When speech stops, gesture stops: Evidence from developmental and crosslinguistic comparisons. *Frontiers in Psychology* 9. 879. https://doi.org/10.3389/fpsyg.2018.00879

Grimm, Hannelore. 2001. *Sprachentwicklungstest für drei- bis fünfjährige Kinder (SETK 3-5): Diagnose von Sprachverarbeitungsfähigkeiten und auditiven Gedächtnisleistungen* [*Language development test for three- to five-year-old children (SETK 3-5): diagnosis of language processing skills and auditory memory skills*]. Göttingen: Hogrefe.

Grimminger, Angela, Katharina J. Rohlfing, Carina Lüke, Ulf Liszkowski, & Ute Ritterfeld. 2019. Decontextualized talk in caregivers' input to 12-month-old children during structured interaction. *Journal of Child Language* 47(2). 418–434. https://doi.org/10.1017/S0305000919000710

Guidetti, Michèle & Elena Nicoladis. 2008. Introduction to special issue: Gestures and communicative development. *First Language* 28(2). 107–115. https://doi.org/10.1177/0142723708088914

Happé, F. G. 1995a. The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development* 66(3). 843–855. https://doi.org/10.1111/j.1467-8624.1995.tb00909.x

Happé, Francesca G. E. 1995b Understanding minds and metaphors: Insights from the study of figurative language in autism. *Metaphor and Symbol* 10(4). 275–295. https://doi.org/10.1207/s15327868ms1004_3

Happé, Francesca G. E. 1994. An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders* 24(2). 129–154. https://doi.org/10.1007/BF02172093

Hayes, Andrew F. 2018. *Introduction to mediation, moderation, and conditional process analysis: Methodology in the social sciences*, 2nd edn. New York: Guilford Press.

Henning, Anne, Tanja Hofer, & Gisa Aschersleben. 2013. *Erweiterte "Theory of Mind-Skala" für 3- bis 11-jährige Kinder* [Extended Theory-of-Mind Scale for 3- to 11-year-old children], 2nd edn. Universität des Saarlandes Arbeitseinheit Entwicklungspsychologie.

Holler, Judith & Stephen C. Levinson. 2019. Multimodal language processing in human communication. *Trends in Cognitive Sciences* 23(8). 639–652. https://doi.org/10.1016/j.tics.2019.05.006

Holyoak, Keith J. & Dušan Stamenković. 2018. Metaphor comprehension: A critical review of theories and evidence. *Psychological Bulletin* 144(6). 641. https://doi.org/10.1037/bul0000145

Hostetter, Autumn B. & Martha W. Alibali. 2019. Gesture as simulated action: Revisiting the framework. *Psychonomic Bulletin & Review* 26. 721–752. https://doi.org/10.3758/s13423-018-1548-0

Huebner, Bryce. 2013. Socially embedded cognition. *Cognitive Systems Research* 25. 13–18. https://doi.org/10.1016/j.cogsys.2013.03.006

Ibáñez, Agustín, Facundo Manes, Josefina Escobar, Natalia Trujillo, Paola Andreucci, & Esteban Hurtado. 2010. Gesture influences the processing of figurative language in non-native speakers: ERP evidence. *Neuroscience Letters* 471(1). 48–52. doi:10.1016/j.neulet.2010.01.009

Iverson, Jana M., Jessie B. Northrup, Nina B. Leezenbaum, Meaghan V. Parladé, Erin A. Koterba, & Kelsey L. West. 2018. Early gesture and vocabulary development in infant siblings of children with autism spectrum disorder. *Journal of Autism and Developmental Disorders* 48(1). 55–71. https://doi.org/10.1007/s10803-017-3297-8

Jafari, Mohieddin & Naser Ansari-Pour. 2019. Why, when and how to adjust your P values?. *Cell Journal (Yakhteh)*, 20 (4). 604–607. https://doi.org/10.22074/cellj.2019.5992

Joue, Gina, Linda Boven, Klaus Willmes, Vito Evola, Liliana R. Demenescu, Julius Hassemer, Irene Mittelberg, Klaus Mathiak, Frank Schneider, & Ute Habel. 2020. Metaphor processing is supramodal semantic processing: The role of the bilateral lateral temporal regions in multimodal communication. *Brain and Language* 205. 104772. doi:10.1016/j.bandl.2020.104772

Kalandadze, Tamar, Courtenay Frazier Norbury, Terje Nærland, & Kari-Anne B. Næss. 2018. Figurative language comprehension in individuals with autism spectrum disorder: A meta-analytic review. *Autism* 22(2). 99–117. https://doi.org/10.1177/1362361316668652

Kauschke, Christina & Julia Siegmüller. 2010. *Patholinguistische diagnostik bei sprachentwicklungsstörungen (PDSS)* [*Diagnosis of developmental language disorders*], 2nd edn. München: Urban & Fischer.

Kelly, Spencer D., Aslı Özyürek, & Eric Maris. 2010. Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science* 21(2). 260–267. https://doi.org/10.1177/0956797609357327

Kendon, Adam. 2004. *Gesture: Visible action as utterance*. Cambridge, UK: Cambridge University Press.

Khatin-Zadeh, Omid. 2023. Embodied metaphor processing: A study of the priming impact of congruent and opposite gestural representations of metaphor schema on metaphor comprehension. *Metaphor and Symbol* 38(1). 70–80. https://doi.org/10.1080/10926488.2022.2122830

Khatin-Zadeh, Omid, Danyal Farsani, Jiehui Hu, Zahra Eskandari, Yanjiao Zhu, & Hassan Banaruee. 2023. A review of studies supporting metaphorical embodiment. *Behavioral Sciences* 13(7). 585. https://doi.org/10.3390/bs13070585

Khatin-Zadeh, Omid, Jiehui Hu, Fernando Marmolejo-Ramos, & Danyal Farsani. 2023. The impact of gestural representation of metaphor schema on metaphor comprehension. *Poznan Studies in Contemporary Linguistics* 59(1). 117–131. https://doi.org/10.1515/psicl-2022-1056

Kita, Sotaro & Aslı Özyürek. 2003. What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language* 48(1). 16–32. doi:10.1016/S0749-596X(02)00505-3

Lecce, Serena, Luca Ronchi, Paola Del Sette, Luca Bischetti, & Valentina Bambini. 2019. Interpreting physical and mental metaphors: Is Theory of Mind associated with pragmatics in middle childhood? *Journal of Child Language* 46(2). 393–407. https://doi.org/10.1017/S030500091800048X

Liszkowski, Ulf, Penny Brown, Tara Callaghan, Akira Takada, & Conny De Vos. 2012. A prelinguistic gestural universal of human communication. *Cognitive Science* 36(4). 698–713. https://doi.org/10.1111/j.1551-6709.2011.01228.x

Liszkowski, Ulf, Malinda Carpenter, & Michael Tomasello. 2007. Reference and attitude in infant pointing. *Journal of Child Language* 34(1). 1–20. https://doi.org/10.1017/S0305000906007689
Liszkowski, Ulf, Malinda Carpenter, & Michael Tomasello. 2008. Twelve-month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *Cognition* 108(3). 732–739. https://doi.org/10.1016/j.cognition.2008.06.013
Liszkowski, Ulf & Michael Tomasello. 2011. Individual differences in social, cognitive, and morphological aspects of infant pointing. *Cognitive Development* 26(1). 16–29. https://doi.org/10.1016/j.cogdev.2010.10.001
Lüke, Carina, Angela Grimminger, Katharina J. Rohlfing, Ulf Liszkowski, & Ute Ritterfeld. 2017. In infants' hands: Identification of preverbal infants at risk for primary language delay. *Child Development* 88(2). 484–492. https://doi.org/10.1111/cdev.12610
Lüke, Carina, Ute Ritterfeld, Angela Grimminger, Ulf Liszkowski, & Katharina J. Rohlfing. 2017. Development of pointing gestures in children with typical and delayed language acquisition. *Journal of Speech, Language, and Hearing Research* 60(11). 3185–3197. https://doi.org/10.1044/2017_JSLHR-L-16-0129
Lüke, Carina, Juliane Leinweber, & Ute Ritterfeld. 2019. Walking, pointing, talking–the predictive value of early walking and pointing behavior for later language skills. *Journal of Child Language* 46(6). 1228–1237. https://doi.org/10.1017/S0305000919000394
Lüke, Carina, Ute Ritterfeld, Angela Grimminger, Katharina J. Rohlfing, & Ulf Liszkowski. 2020. Integrated communication system: Gesture and language acquisition in typically developing children and children with LD and DLD. *Frontiers in Psychology* 11. 118. https://doi.org/10.3389/fpsyg.2020.00118
Mainela-Arnold, Elina, Martha W. Alibali, Autumn B. Hostetter, & Julia L. Evans. 2014. Gesture–speech integration in children with specific language impairment. *International Journal of Language & Communication Disorders* 49(6). 761–770. https://doi.org/10.1111/1460-6984.12115
McGillion, Michelle, Jane S. Herbert, Julian Pine, Marilyn Vihman, Rory dePaolis, Tamar Keren-Portnoy, & Danielle Matthews. 2017. What paves the way to conventional language? The predictive value of babble, pointing, and socioeconomic status. *Child Development*, 88(1), 156–166. https://doi.org/10.1111/cdev.12671
McNeill, David. 2014. The emblem as metaphor. In M. Seyfeddinipur & M. Gullberg (eds.), *From gesture in conversation to visible action as utterance*, 75–94. John Benjamins. https://doi.org/10.1075/z.188.05nei
McNeill, David. 1992. *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.
Milligan, Karen, Janet Wilde Astington, & Lisa Ain Dack. 2007. Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Development* 78(2). 622–646. https://doi.org/10.1111/j.1467-8624.2007.01018.x
Norbury, Courtenay Frazier. 2005. The relationship between theory of mind and metaphor: Evidence from children with language impairment and autistic spectrum disorder. *British Journal of Developmental Psychology* 23(3). 383–399. https://doi.org/10.1348/026151005X26732
Osterhaus, Christopher & Susanne Koerber. 2021. The development of advanced theory of mind in middle childhood: A longitudinal study from age 5 to 10 years. *Child Development* 92(5). 1872–1888. https://doi.org/10.1111/cdev.13627
Osterhaus, Christopher, Susanne Kristen-Antonow, Daniela Kloo, & Beate Sodian. 2022. Advanced scaling and modeling of children's theory of mind competencies: Longitudinal findings in 4-to 6-year-olds. *International Journal of Behavioral Development* 46(3). 251–259. https://doi.org/10.1177/01650254221077334
Özçalışkan, Şeyda. 2005. On learning to draw the distinction between physical and metaphorical motion: Is metaphor an early emerging cognitive and linguistic capacity? *Journal of Child Language* 32(2). 291. https://doi.org/10.1017/S0305000905006884
Özyürek, Aslı. 2014. Hearing and seeing meaning in speech and gesture: Insights from brain and behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369(1651). 20130296. https://doi.org/10.1098/rstb.2013.0296
Perner, Josef & Heinz Wimmer. 1985. "John thinks that Mary thinks that…" attribution of second-order beliefs by 5-to 10-year-old children. *Journal of Experimental Child Psychology* 39(3). 437–471. https://doi.org/10.1016/0022-0965(85)90051-7
Peterson, Candida C., Henry M. Wellman, & Virginia Slaughter. 2012. The mind behind the message: Advancing theory-of-mind scales for typically developing children, and those with deafness, autism, or Asperger syndrome. *Child Development* 83(2). 469–485. doi:10.1111/j.1467-8624.2011.01728.x
Pouscoulous, Nausicaa & Michael Tomasello. 2020. Early birds: Metaphor understanding in 3-year-olds. *Journal of Pragmatics*, 156, 160–167. https://doi.org/10.1016/j.pragma.2019.05.021
Pronina, Mariia, Pilar Prieto, Luca Bischetti, & Valentina Bambini. 2023. Expressive pragmatics and prosody in young preschoolers are more closely related to structural language than to mentalizing. *Language Learning and Development* 19(3). 323–344. https://doi.org/10.1080/15475441.2022.2074852
Rakoczy, Hannes, Raphaela Wandt, Stefanie Thomas, Jana Nowak, & Ute Kunzmann. 2018. Theory of mind and wisdom: The development of different forms of perspective-taking in late adulthood. *British Journal of Psychology* 109 (1). 6–24. https://doi.org/10.1111/bjop.12246
Rakoczy, H., Antje Harder-Kasten, & Lioba Sturm. 2012. The decline of theory of mind in old age is (partly) mediated by developmental changes in domain-general abilities. *British Journal of Psychology* 103(1). 58–72. https://doi.org/10.1111/j.2044-8295.2011.02040.x

Rohlfing, Katharina J., Carina Lüke, Ulf Liszkowski, Ute Ritterfeld, & Angela Grimminger. 2022. Developmental paths of pointing for various motives in infants with and without language delay. *International Journal of Environmental Research and Public Health* 19(9). 4982. https://doi.org/10.3390/ijerph19094982

Rowe, Meredith L. & Susan Goldin-Meadow. 2009. Differences in early gesture explain SES disparities in child vocabulary size at school entry. *Science* 323(5916). 951–953. https://doi.org/10.1126/science.1167025

Schaunig, Ines, Ulrike Willinger, & Anton K. Formann. 2004. Das Verständnis metaphorischer Sprache bei Grundschulkindern [Children's Comprehension of Metaphoric Language]. *Zeitschrift für Pädagogische Psychologie*, 18(1). 53–61. https://doi.org/10.1024/1010-0652.18.1.53

Schnell, Zsuzsanna. 2007. Metaphor processing and the acquisition of idioms: A mentalistic model. *Acta Linguistica Hungarica* 54(1). 73–104. https://doi.org/10.1556/aling.54.2007.1.3

Sodian, Beate & Susanne Kristen-Antonow. 2015. Declarative joint attention as a foundation of theory of mind. *Developmental Psychology* 51(9). 1190–1200. https://doi.org/10.1037/dev0000039

Tellegen, Peter J., Jacob A. Laros, & Franz Petermann. 2007. *SON-R 2 1/2–7: Non-verbaler intelligenztest*. Göttingen: Hogrefe.

Tonini, Elisabetta, Luca Bischetti, Paola Del Sette, Eleonora Tosi, Serena Lecce, & Valentina Bambini. 2023. The relationship between metaphor skills and Theory of Mind in middle childhood: Task and developmental effects. *Cognition* 238, 105504. https://doi.org/10.1016/j.cognition.2023.105504

Vogt, Swetlana & Peter Indefrey. 2017. Metaphernerwerb: eine empirische Studie bei Kindern im Alter von sechs bis vierzehn Jahren [Metaphor acquisition: an empirical study with children between the ages of six to fourteen years]. *Metaphorik.de*, 27/2017. 69–106.

Vulchanova, Mila, Evelyn Milburn, Valentin Vulchanov, & Giosuè Baggio. 2019. Boon or burden? The role of compositional meaning in figurative language processing and acquisition. *Journal of Logic, Language and Information* 28(2). 359–387. https://doi.org/10.1007/s10849-019-09282-7

Vulchanova, Mila, David Saldaña, Sobh Chahboun, & Valentin Vulchanov. 2015. Figurative language processing in atypical populations: the ASD perspective. *Frontiers in Human Neuroscience* 9. 24. https://doi.org/10.3389/fnhum.2015.00024

Wellman, Henry M., David Cross, & Julanne Watson. 2001. Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development* 72(3). 655–684. doi:10.1111/1467-8624.00304

Wellman, Henry M. & David Liu. 2004. Scaling of theory-of-mind tasks. *Child Development* 75(2). 523–541. https://doi.org/10.1111/j.1467-8624.2004.00691.x

Wimmer, Heinz & Josef Perner. 1983. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13(1). 103–128. https://doi.org/10.1016/0010-0277(83)90004-5

White, Sarah, Elisabeth Hill, Francesca Happé, & Uta Frith. 2009. Revisiting the strange stories: Revealing mentalizing impairments in autism. *Child Development* 80(4). 1097–1117. https://doi.org/10.1111/j.1467-8624.2009.01319.x

Wray, Charlotte, Courtenay Frazier Norbury, & Katie Alcock. 2016. Gestural abilities of children with specific language impairment. *International Journal of Language & Communication Disorders* 51(2). 174–182. https://doi.org/10.1111/1460-6984.12196

Zhao, Xinshu, John G. Lynch, Jr., & Qimei Chen. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research* 37(2). 197–206. https://doi.org/10.1086/651257

Appendix A

Summary of Vogt & Indefrey's (2017) study on metaphor acquisition

Vogt & Indefrey (2017) conducted their metaphor acquisition study with 60 children between the ages of 6-14 years, separated into three age groups spanning: 6-7 years (*n* = 20), 9-11 years (*n* = 20), and 12-14 years (*n* = 20), who all spoke German as a first language. The metaphors were originally taken from 2 corpora (DWDS (digitales Wörterbuch der deutschen Gegenwartssprache [digital dictionary of contemporary German]) and COSMAS II (Corpusdatenbank des Instituts für deutsche Sprache in Mannheim [Corpus database of the Institute for the German Language in Mannheim]), as well as the "Deutscher Wortschatz der Universität Leipzig" (German vocabulary from the University of Leipzig) and from various reference books (see Vogt & Indefrey 2017 for further citations). They selected only the metaphors that contained words whose denotative meanings would usually already be a part of the vocabulary of children about to start school (Augst 1984, as cited in Vogt & Indefrey, 2017). These metaphors were then pilot-tested for general comprehensibility by 60 speakers of German as a first language and then reduced to those which were most widely recognized by native speakers. The authors created a taxonomy of metaphors, establishing 11 (sub-)categories: 1) techno- and biomorphic metaphors, referring to the process of assigning specific visual or functional properties from the source domain (e.g., objects/artefacts, food, or plants) to living organisms. 2) Anthropomorphic metaphors, referring to the attribution of human characteristics to inanimate objects. 3) Animal metaphors, referring to the two-way mapping between [HUMAN – ANIMAL – HUMAN] where human characteristics or behaviors are first assigned to animals by means of transfer and then supposedly "animal" characteristics are assigned to humans by means of retransfer or re-predication. 4) Synesthetic metaphors, which typically make use of adjectives referring to different sensory qualities (or qualia) and represent a special form of processing and linking of multisensory processes, these include: 4.1) synesthetic metaphors of a perceptual character and 4.2) synesthetic metaphors of an emotional character, in which all elements are derived from perceptual areas, in contrast to 4.3) pseudo-synesthetic metaphors, in which the modifying element derives from the perceptual while the modified element comes from a non-perceptual area. 5) Spatial metaphors, which involve orientation or pictorial schematics and consist of a transfer from their central component term denoting "spatial form" or "figurativeness" to an abstract area. 6) Metaphors that refer to the type of action conveyed by the sentential verb and are mostly used with reference to abstract entities, often involving the personification of both concrete and abstract units, the depersonalization of animated objects, and a change in the semantic role of the verb: 6.1) process metaphors, 6.2) action metaphors, 6.3) action metaphors relating to acts or actions that express an emotional state, and 6.4) condition metaphors relating to emotional states (Vogt & Indefrey 2017).

They found between-group differences concerning the ages at which children acquire different categories of metaphor which we have summarized within a table format below.

Table A1. Summary of Vogt & Indefrey's (2017) findings concerning children's developing understanding of metaphor categories.[a]

Metaphor Type	Age Group 1 (6-7 years old) (%)	Age Group 2 (9-11 years old) (%)	Age Group 3 (12–14 years old) (%)
Technomorphic	20	57	67
Anthropomorphic	60	92	99
Animal	62	85	92
Synesthetic (Perceptual Character)	43	63	80
Synesthetic (Emotional Character)	55	85	91
Pseudo-Synesthetic	33	58	85
Spatial	19	49	90
Process Metaphors	33	73	91
Action Metaphors	53	86	94
Action Metaphors (Emotional States)	28	55	89
Condition Metaphors (Emotional States)	37	74	92

[a]Averaged proportion of children who demonstrated accurate understanding of the metaphors within each category.

Appendix B

Minor adaptations of the stimuli to the digital or modern context

ToM-Scale. The pictures used for this measure had to be adapted from their original physical version for digital use. The original images from the authorized German-language translation of the Extended Theory-of-Mind Scale (originally by Peterson et al. 2012; translated by Henning et al. 2013) were displayed using the screen-sharing function of the digital conferencing tool. In the original version of the hidden emotion subtest, two pictures are laid out on the table in front of the child. As this could not be easily achieved in a digital setting, we incorporated them to make one image for the screen where the picture of the character in the story was centered and positioned above the picture of the facial expressions (happy, in-between, and sad). In the original task the children would be asked to point to the relevant facial expression when answering questions about the story. To provide clarity for their responses, we therefore adapted the image to have different and easily distinguishable colored outlines around each of the three facial expressions, and the children were asked to clearly identify the relevant expression by the color of the outline box around it. For the sarcasm subtest, the image was simply shown on the screen as it would have been laid out on the table in front of the child. Where an experimenter would point to the characters in an in-person setting, a spotlight cursor was used on the screen instead. In the original version of the explicit false belief subtest a small toy figurine of a boy would be placed on the table near the picture but in the middle of the images of the two locations in the story. We therefore adapted the original image to have a cartoon picture of the boy (represented originally by the toy figurine) positioned at the bottom of the screen equidistantly between the two location images.

Strange Stories. We updated two of the eight German translations (Rakoczy et al. 2012, 2018; Ebert 2020) of the Strange Stories to be more representative of contemporary children's experience. These did not change the target meaning of the story and were only very minor changes. One of the persuasion stories was changed so that the protagonist attempted to obtain pieces of pizza rather than mini-sausages. One of the white-lie stories was also changed so that the undesired Christmas present was a pile of books rather than an old brand of an encyclopedia set that children today might not be familiar with.

Appendix B: Publication II

Crawshaw, C. E., Lüke, C., & Ritterfeld, U. (2025). Late Talkers' Language, Metaphor, Theory of Mind, and Reading Skills at 9 Years. *Journal of Speech, Language, and Hearing Research*, *68*(3), 1038–1055. https://doi.org/10.1044/2024_JSLHR-23-00703

Since only the accepted manuscript could be included in this thesis, please note that a missing reference was added during the production of the publisher's version, which could not be appended here for copyright reasons. This reference is instead included in section 8: References, which should be regarded as the definitive references list of this dissertation and its appendices.

Late Talkers' Language, Metaphor, Theory of Mind, and Reading Skills at 9 Years

Author List and Affiliations:

Camilla E. Crawshaw*[1], Carina Lüke[2], Ute Ritterfeld[1]

[1] Language and Communication, Department of Rehabilitation Sciences, TU Dortmund University, Dortmund, Germany

[2] Special Education and Therapy in Language and Communication Disorders, Institute of Special Education, Faculty of Human Sciences, Julius-Maximilians-University of Würzburg, Würzburg, Germany

*Corresponding author (email address: camilla.crawshaw@tu-dortmund.de)

Conflict of interest statement: All authors confirm that they hold no conflicts of interest and that they do not stand to gain personally or financially from the subject matter or materials discussed in this manuscript.

Abstract

Purpose: Prior work has found that 'late talkers' (LTs) as a group continue to demonstrate lower language and reading outcomes compared to their typically developing (TD) peers even into young adulthood. Others identified that children diagnosed with developmental language disorder (DLD) show difficulties later with Theory of Mind (ToM) tasks and metaphor comprehension but there is a shortage of research specifically investigating these advanced skills in LTs. The current study therefore compared language-related skills of former LTs with their TD peers at school age.

Method: A longitudinal sample (N = 35) of monolingual German-speaking children was observed from age 1;0 until 9;0, comprising TD children (n = 27), and children identified as LTs at age 2;0 (n = 8) of which two met criteria for DLD between ages 3;0–6;0. Children's language (productive vocabulary, productive and receptive grammar), reading, metaphor comprehension, and theory of mind skills (ToM-scale and Strange Stories) were investigated and group comparisons were conducted.

Results: Former LTs performed worse than the TD children on measures of productive vocabulary, receptive grammar, metaphor comprehension, and the ToM Strange Stories task at the age of 9;0 but not on measures of productive grammar, reading or the ToM-scale.

Conclusions: The findings indicate that LTs can catch up with their TD peers in some areas of language and ToM but that subtle differences remain across other complex areas. Further research is needed to pinpoint possible explanations for why certain skills are more strongly impacted and the potential developmental interactions between these competencies.

Keywords: Late Talkers; ToM; Longitudinal Study

Developmental Paths among Late Talkers

In contrast to children who are typically developing (TD) or who go on to manifest a developmental language disorder (DLD), less focus has been placed on the outcomes of ‘late talkers’ (LTs) who seem to ‘catch up’ with their TD peers, especially regarding their abilities beyond structural language. LT status refers to children who have no known cognitive impairment or hearing loss, demonstrate a comparably small productive vocabulary between the ages of 12–24 months, do not start to combine words at the age of 24 months, and demonstrate differences in both noun and verb acquisition (c.f. Horvath et al., 2022; Perry et al., 2023; Rescorla, 2009, 2011; Sansavini et al., 2021). Other differences can persist in speech processing and visual attention (Perry et al., 2023). Estimates of prevalence vary between 10–20% of all two-year-olds (c.f. Zubrick et al., 2007). Many LTs go on to reach expectations set for their TD peers although for some this may not be until they are 5 years old, and up to 40% of LTs will not, potentially going on to receive a diagnosis of DLD (e.g. Bates et al., 1995; Bishop, 2017; Norbury et al., 2016; Rescorla, 2011). At first, those children who do appear to catch up were believed to resume a typical developmental trajectory. However, more recent work has observed that they experience persistently lower outcomes in language and language-related skills, even into early adulthood (c.f. Bates et al., 1995; Horvath et al., 2022; Perry et al., 2023; Rescorla, 2009, 2011; Sansavini et al., 2021). This suggests that TD children, LTs and children with DLD may fall—albeit differing quantitatively from one another—along dimensional spectra of language(-related) abilities (Rescorla, 2009, 2011; Thal et al., 2013).

Early findings (e.g. Bates et al., 1995; Rescorla, 1997) were published more than 25 years ago. Yet since then relatively few studies have analysed the development of language and reading alongside other language-related skills (e.g. Theory of Mind or figurative language) in older children with a history of being LTs but who do not necessarily meet criteria for DLD. We will investigate where LTs might fall compared to their peers along linguistic (productive vocabulary, productive grammar, receptive grammar) and language-related developmental spectra at school age. This includes reading comprehension and two other complex skills bridging language and social cognition: Theory of Mind (ToM) and metaphor comprehension. With reference to prior work, the following introductory sections will briefly address each skill and present the current picture of their development in LTs. Where gaps exist, we will consider those with DLD for analogical potential.

Language

A number of longitudinal studies have followed LTs until the ages of 4, 5 or 6 (e.g. Hammer et al., 2017) but only a scarce few have continued beyond (for a discussion, see Rescorla, 2009). One longitudinal cohort of 56 LTs (Bishop & Edmundson, 1987) were subdivided at age 5;6 into those whose language issues had resolved (n = 26) versus those whose had not (n = 30) and followed up until the age of 15 (Stothard et al. 1998). Compared to age-matched TD controls, the children with resolved language issues did not differ significantly on vocabulary and language comprehension skill tests but did perform significantly less well on phonological processing and literacy skill tests (Stothard et al., 1998). The children with persistent language issues demonstrated significantly lower performance across all measured areas of spoken and written language (Stothard et al., 1998). Rescorla (2009) followed the development of 26 children identified as LTs at intake (24–31 months) who had typical nonverbal cognitive ability and receptive language, alongside 23 TD children matched at intake on age, socioeconomic status, and nonverbal ability, until they were 17 years old. Rescorla's (2009) LTs demonstrated performance in the average range on all language and reading tasks at age 17 but still achieved significantly lower vocabulary, grammar, and verbal memory scores than their TD peers, despite no significant differences in their reading and writing scores. Thus, evidence from prior work indicates that LTs' performance on language and reading measures might change over time, differ at both the individual and sample level, and depend upon how these skills are measured. Even so, it has also revealed that many LTs do experience long-lasting, residual challenges in their language outcomes.

Reading

Many studies have already demonstrated that children's early oral language skills can predict their later reading comprehension abilities (e.g. LARRC & Chiu, 2018; Hjetland et al., 2019). However, a distinction between LT and TD reading abilities seems less clear. In Bishop, Edmundson, Stothard and colleagues' (1987; 1998) longitudinal sample, LTs performed significantly worse at age 15 than their TD peers on reading skills comprising single word reading, single word spelling, and reading comprehension. In Rescorla's (2009) sample at age 17, however, there were no significant differences across a set of subtests evaluating ability to decode words on a list, timed reading and comprehension of short statements, and timed writing of short statements using target words. Interestingly, when Bishop and Edmundson's (1987) sample were assessed at age 8;6, the 'recovered' LTs (non-DLD) displayed no reading or spelling difficulties and performed within the normal range on tasks tapping phonological strategies: non-word reading and spelling (c.f. Stothard et al., 1998). In contrast, although Rescorla's (2009) sample mostly performed within the average range on reading measures when they were aged 8, 9, and 13, they still demonstrated significantly lower performance than their TD peers. These inconsistencies might suggest that longitudinal findings regarding LTs are influenced by differences at both the group- and the individual level within children's language and literacy development (Bates et al., 1995).

More recent, fine-grained approaches to literacy research may support this lack of a clear-cut distinction in LTs' comparative reading skills. A longitudinal study by Psyridou et al. (2018) with 200 children aged 2–16 years found that LTs who developed dyslexia had experienced both expressive and receptive vocabulary delay as well as a family risk for dyslexia. These children struggled with reading comprehension but not reading fluency, sustaining difficulties into adolescence, while LTs without receptive vocabulary difficulties generally became typical readers (Psyridou et al., 2018). Psyridou et al. (2018) argued that being a LT was not a sufficient risk index for developing reading comprehension difficulties. This would align with prior work, signifying different types of and severities within LTs (e.g. Thal et al., 2013) which do not facilitate simple binary classifications (Dollaghan, 2013), and highlighting a potential need for the building of subgroups.

Theory of Mind (ToM)

ToM refers to the ability to understand and attribute one's own and others' mental states and representations, taps perspective-taking skills, and is closely linked to other linguistic, pragmatic, cognitive and social skills (c.f. Devine & Lecce, 2021). Within the TD population, previous work has already established that language and ToM are co-developing skills, with earlier language skills predicting later ToM performance (for a meta-analysis, see Milligan et al. 2007). However, researchers are still divided on exactly which components of language are facilitative. Some authors have specified grammatical constructions such as sentential complementation (e.g. de Villiers, 2007; Durrleman et al., 2017). Some have highlighted the impact of receptive grammar and sentence comprehension skills (Ebert, 2020; De Mulder et al., 2019). Others have instead emphasised the role of more general language and communicative skills (e.g. Milligan et al., 2007; Ensor et al., 2014). It has been demonstrated that vocabulary can also independently predict ToM performance (De

Mulder et al., 2019; Devine et al., 2016; Ebert, 2020; Happé, 1995; de Villiers, 2005) and some have underlined the importance of mental state verbs (Astington & Baird, 2005).

ToM in LTs

To our knowledge, no existing work has gone beyond expressive and receptive language or reading skills to consider metacognitive skills in LTs who do not necessarily go on to manifest DLD. This gap should be addressed since extensive prior work has shown early linguistic skills are formed by and chronologically aligned with the emergence of non-linguistic skills (c.f. Thal et al., 2013). Thal et al.'s (2013) longitudinal study followed LTs until the age of 7, identifying a spectrum spanning TD to LT with delayed production but typical comprehension (late producer) and on further to LT with both delayed production and comprehension (late comprehender). Group differences were evident at as early as 10 months of age and across this spectrum, representational skills (in terms of gesture use) were shown to be progressively poorer and lowest in late comprehenders (Thal et al., 2013). Compared to their TD peers, LTs appear to engage in less symbolic play, a precursor to ToM (Paul & Ellis Weismer, 2013). LTs followed over a 15-year longitudinal study used fewer cognitive mental state terms (e.g. "think", "know") at age 5 than their TD peers (Rescorla, 2013); these terms are important for the linguistic embedding of ToM concepts. Early differences in representational ability could be associated with subsequent delays in ToM development (a metarepresentational ability). Alongside TD or autistic populations, prior work has only addressed ToM skills in LTs who received a diagnosis of DLD. Since LTs who appear to have 'recovered' likely fall somewhere between their TD and DLD peers along dimensional spectra of performance (Rescorla, 2009, 2011; Thal et al., 2013), we will next discuss existing work on the DLD population to provide a situative context.

ToM in Children with DLD

Children with DLD face issues with language production, comprehension, and processing. They experience varying degrees of impairment across broad-ranging areas of language, including lexical, morpho-syntactic, pragmatic, as well as both oral and written skills (Sansavini et al., 2021). Children with DLD often have long-term learning difficulties and can struggle with behavioural, psychological, emotional, and social adaptation, affecting their ability to work and form relationships in adulthood (for a recent review of predictors and outcomes for DLD, see Sansavini et al., 2021). On the whole, prior work has shown that ToM skills are also a delayed area in DLD (Andrés-Roqueta et al., 2013; Durrleman & Delage, 2020; Durrleman et al., 2017; Farrant, 2015; Farrant et al., 2006; Farrar et al., 2009; Gillott et al., 2004; Rakhlin et al., 2011; Spanoudis, 2016; Nilsson & de Lopez, 2016; Vissers & Koolen, 2016; Smit et al., 2019). ToM ability covers many different aspects of mentalising skills and a variety of tasks exist to test it (for a recent discussion and overview, see Devine & Lecce, 2021). It is therefore plausible that the type of task utilised might influence whether researchers identify differences in performance across TD children, those who are LTs, and those with DLD. In Table 1 we present an overview of recent studies addressing ToM performance in children with DLD. As can be seen from the table, the studies vary widely in terms of participant ages and ToM tasks used.

Table 1: *Overview of recent studies considering ToM performance in children with DLD*.

In a meta-analysis of ToM skills in DLD, Nilsson and de Lopez (2016) evaluated 17 studies covering a total of 745 children between ages 4–12 and found that children with DLD performed substantially worse on ToM tasks than their age-matched TD peers. Their findings reinforce the idea that (early) language and ToM skills are associated, with language facilitating age-appropriate ToM development. They also indicate a potential interface, with impairment in one domain extending into the other (Nilsson & de Lopez, 2016). Vissers and Koolen (2016) conducted a review of studies investigating social-emotional functioning and ToM abilities in children with DLD aged 2;3–6;2. They concluded that preschoolers with DLD experienced social-emotional difficulties and impairments in both cognitive (imitation, joint attention, and false belief understanding) and affective (recognising and understanding emotions) ToM (Vissers & Koolen, 2016). Smit et al. (2019) reviewed studies on social emotional and ToM ability in adolescents (ages 10–24) with DLD or who were deaf or hard of hearing, examining parallels and establishing a framework that was mediated by limited linguistic competence or restricted language exposure.

Metaphor Comprehension

Metaphor comprehension requires making a non-literal mapping between concepts and linguistic forms to access and understand an interlocutor's intended meaning. This is often ambiguous and may require perspective-taking skills. Metaphor comprehension has thus been linked with both semantic skills and ToM ability (Deckert et al., 2019; Kalandadze et al., 2018; Lecce et al., 2019; Norbury, 2005). This indicates that metaphor might function as a form of interface between ToM and language (c.f. Pronina et al., 2023). Bidirectional longitudinal associations have been found between TD 9-year-old children's ability to understand metaphors and their peer acceptance or rejection outcomes (Del Sette et al., 2021). Work highlighting ToM as a predictor of social functioning ability has also suggested that adolescents with DLD's difficulties forming peer relationships stem from their pragmatic language impairments (Smit et al., 2019). These findings underline that the ability to comprehend metaphors is not merely a useful academic skill but instead crucially important for children's social outcomes. As a pragmatic, metalinguistic skill interfacing ToM and language abilities, metaphor comprehension may thus be especially relevant to consider in LTs. It could also provide a potential opportunity to narrow down challenging developmental domains and help pinpoint underlying, causal deficits. Building on the work of Thal et al. (2013), if some (late comprehender) but not all (late producer) LTs struggle with early representational skills, then metaphor comprehension might be a useful domain to help build and distinguish LT subgroups. Since no prior work appears to have considered metaphor comprehension in LTs without DLD, next we present the findings regarding the DLD population.

Metaphor Comprehension in DLD

In contrast to ToM, only very few studies have considered how well children with DLD comprehend and process metaphors or other types of figurative or abstract language (e.g. Bühler et al., 2018; Lorusso et al., 2015; Norbury, 2005; Spanoudis, 2016). Those addressing metaphor comprehension also incorporated measures of ToM. Norbury (2005) compared TD children (*n* = 34), language-impaired children (*n* = 28), language-impaired autistic children (*n* = 31), and non-language-impaired autistic children (*n* = 29) across semantic knowledge, ToM, and metaphor comprehension measures. Results showed specifically that both groups of children with language impairment (with and without autistic status) performed significantly worse across all three measures (Norbury, 2005). However, this 'language impairment' classification comprised both children with DLD and another language disorder associated with autism. Spanoudis (2016) found that children with DLD performed significantly worse than TD controls on novel metaphor and simile comprehension. Bühler et al. (2018) investigated novel metaphor comprehension in children with DLD (*n* = 15) aged 3;6–4;1 compared to age-matched TD peers (*n* = 15) and language-matched younger TD children (*n* = 15). Children with DLD performed less well than age-matched controls but similarly to language-matched controls, suggesting their delay in metaphor comprehension arose from general, overall linguistic competence rather than difficulties in pragmatic inference making (Bühler et al., 2018). Both Norbury (2005) and Spanoudis (2016) found correlations between language, ToM, and metaphor comprehension abilities. However, given (a) the paucity of studies on metaphor comprehension in DLD, (b) work suggesting that metaphor comprehension and ToM are linked, and (c) findings indicating that ToM skills are impaired or at least delayed in DLD, it remains difficult to conclude whether children with DLD's metaphor comprehension struggles are purely caused by linguistic issues or whether there may be other contributing factors.

Addressing Potential (Meta-)Representational, Inferential, and Decoding Deficits in LTs

Children develop an understanding of pictorial and linguistic representation—enabling them to acquire vocabulary—as a precursor to understanding the metarepresentation—representing one's own and others' representations—required for explicit ToM (Doherty, 2008). Being able to read also requires understanding that written letters, words, and text refer to and represent some object, concept, condition, or context in the real or a possible world. Thus, reading, ToM, and metaphor comprehension may all tap semantic, inferential, or decoding skills in potentially analogous ways. Moreover, to understand written narratives, children actively require ToM skills to attribute mental states, thoughts, and emotions to characters (Gordon Pershey, 2000). Some prior research had identified direct connections between early and advanced ToM skills and later reading comprehension (Atkinson et al., 2017; Boerma et al., 2017) but more recent longitudinal work has indicated this connection is largely mediated by early language skills (Ebert, 2020). On a potential spectrum or even complex continuum of interrelated competencies, reading ability and figurative language comprehension might arguably interact more closely. In a study with 199 TD children, Levorato et al. (2004) found that the ability to understand a text predicted children's understanding of idioms in context. In sum, connections between reading comprehension and ToM or metaphor comprehension may be complex, draw on similar underlying skills, and may not be easy to disentangle. However, early language skills clearly impact these later skills and their development should therefore be addressed in LTs as well. ToM is an example of a metarepresentational, metacognitive ability while metaphor comprehension is a metalinguistic skill that also taps representational abilities. Investigating these skills might reveal important associations between very early and later complex language development in LTs.

Research Aim

Prior work (e.g. Rescorla, 2002, 2009; Rescorla et al., 1997; Stothard et al., 1998) has found that LTs with and without a diagnosis of DLD continue to demonstrate differences from their TD peers in language outcomes even into adolescence. Findings are mixed regarding reading-related skills; these may be impacted at different developmental stages (Bishop & Edmundson, 1987; Stothard et al. 1998; Rescorla, 2009). More recent work suggests reading problems may only occur in combination with a family risk for dyslexia (Psyridou et al., 2018). Both ToM and metaphor comprehension seem delayed in children with DLD (e.g. Nilsson & de Lopez, 2016; Norbury, 2005) but no study has yet investigated whether these skills are also impacted in LTs who appear to 'recover'.

Earlier language ability predicts later performance on ToM (e.g. Milligan et al., 2007) as well as metaphor processing tasks (e.g. Deckert et al., 2019; Kalandadze et al., 2018) and ToM has been associated with metaphor comprehension (e.g. Lecce et al., 2019; Norbury, 2005). If language ability is conceptualised as a spectrum ("language endowment spectrum", cf. Ellis Weismer, 2007, p.84), then it is likely that LTs would also fall somewhere between TD and DLD children along spectra or continua of language-associated skills such as ToM and metaphor comprehension. The current study thus sought to explore 2 research questions: (1) Are language and reading abilities impacted in (German-speaking) LTs at age 9;0? (2) Are metacognitive and metalinguistic (ToM and metaphor comprehension) abilities affected in 9;0-year-old children with a history of being a LT?

Method

Participants

The children participating in this study were 35 (18 boys, 17 girls) monolingual German-speaking children who had previously taken part in a wider longitudinal study between the ages of 1;0–6;0 years (for further details, see Lüke et al., 2020; Crawshaw et al., submitted). To address the current research goals we worked with the children at age 9;0 (age at first testing session: *M* = 9 years 0 months 16 days, *SD* = 11 days, age at second session approximately a week later: *M* = 9 years 0 months 24 days, *SD* = 12 days, age at third session approximately a month after first testing: *M* = 9 years 1 month 18 days, *SD* = 31 days). Of the 35 children, 8 were characterised as LTs at age 2;0 (4 boys, 4 girls). Of these 8, 2 (both girls) received a diagnosis of DLD between ages 3;0–6;0 but no longer met diagnostic criteria at age 9;0 so we decided to keep them under the grouping of LTs.

Criteria for LT

The criteria for definition as a LT (at age 2;0) or as having DLD (from ages 3;0–6;0) stemmed from the children’s scores on standardised language measures and information received via a parental report questionnaire. At age 2;0 this consisted of seven measures:

- FRAKIS: Fragebogen zur frühkindlichen Sprachentwicklung [Questionnaire for Early Language Acquisition], German MCDI-version for children between 18 and 30 months including the subscales *vocabulary size*, *morphological skills*, and *syntactic skills* (Szagun et al., 2009).
- SETK-2: Sprachentwicklungstest für zweijährige Kinder [Language Acquisition Test for 2-year-old Children], including the subtests *word comprehension*, *sentence comprehension*, *word production*, and *sentence production* (Grimm, 2000).

At age 3;0, three measures were used:

- SETK 3-5: Sprachentwicklungstest für drei- bis fünfjährige Kinder [Language Acquisition Test for 3- to 5-year-old Children], including the subtests *sentence comprehension* and *sentence production* (Grimm, 2001).
- PDSS: Patholinguistische Diagnostik bei Sprachentwicklungsstörungen [Patholinguistic Diagnostics for Developmental Language Disorder], containing the subtest *word production*, (Kauschke and Siegmüller, 2010).

At age 4;0, two measures were used:

- Potsdam-Illinois Test für Psycholinguistische Fähigkeiten [Potsdam-Illinois Test of Psycholinguistic Abilities], including the subtests *word production* and *grammar production*, (Esser and Wyschkon, 2010).

At ages 5;0 and 6;0, three measures were used:

- the same two tests as at age 4;0; and
- the TROG-D: Test zur Überprüfung des Grammatikverständnisses [German version of the Test for Reception of Grammar] (Fox, 2013).

To meet criteria for LT status, children had to score results in at least one language subtest that were 1½ SD below the mean (i.e. a T-score of ≤35) and 1 SD below the mean in at least one further language subtest (i.e. a T-score of <40). These diagnostic criteria were defined in the early stages of this study, many years prior to the work of the CATALISE consensus across English-speaking countries (Bishop et al., 2017) and are more aligned with the results of a Delphi study later conducted across German-speaking countries (Lüke et al., 2023).

At age 3;6 children's nonverbal IQ was assessed using the SON-R (Tellegen et al., 2007), indicating no differences between LT (*Mdn* = 98.0, *IQR* = 25.0) and TD (*Mdn* = 108.0, *IQR* = 12.0, U = 56.0, Z = 1.44, p = 0.151) groups.

Via a questionnaire at age 9;0, we asked parents to inform us of any diagnoses their children had received since the previous round of data collection at age 6;0. In the intervening three years, four of the children (3 TD, 1 LT) had been officially diagnosed with dyslexia. The task used in this study to measure reading performance would not be appropriate for confirming or issuing a diagnosis of dyslexia (prevalent in the LT population). Since we did not systematically assess dyslexia, confirm any outside diagnoses or seek potential new ones, we have not addressed it further in our analyses. However, parents were informed and advised that they may wish to seek further support for their children when their performance had been lower than might be expected.

Participants' socioeconomic status as well as their demographic and family background data have already been published in full elsewhere (cf. Lüke, Grimminger et al., 2017) but they came from a WEIRD (Western, Educated, Industrialised, Rich, and Democratic) society. Before commencing the study, all procedures, measures, and assessment of participants were evaluated for ethical considerations and granted approval by the Internal Review Board of TU Dortmund University (reference code: GEKTUDO-2020-13). Informed consent was sought from the parents or caregivers of the participating children, and they were able to withdraw their child along with all relevant data from the study at any time. As each testing session began, the children were also asked to verbally confirm that they wanted to participate. They were informed that they could withdraw from the testing sessions whenever they wished at no disadvantage to themselves or their families.

Procedure

Each child took part in three testing sessions: the first two sessions were conducted online using a digital conferencing tool (Zoom) and the final session was conducted in-person within our lab. During the first session we utilised a measure of productive vocabulary and the ToM-scale. In the second session we employed a metaphor comprehension task, a reading task, and the Strange Stories tasks. During the third session we executed measures of productive and receptive grammar. Two research assistants, oblivious to the specific research questions, carried out the testing sessions. Each child consistently interacted with the same experimenter across the three sessions, except for one child who experienced the first two online sessions with one experimenter and the final in-person session with the other experimenter. This scheduling conflict resulted from an unfortunately unavoidable delay caused by the ongoing COVID-19 pandemic. During the online sessions, each child was tested from their home using a laptop or tablet which belonged to their family. All experimental interactions were audio- and video-recorded.

Measures

Productive Vocabulary

We measured the children's vocabulary skills using the productive vocabulary subtest from the standardised German-language version of the (Potsdam-)Illinois Test for Psycholinguistic Abilities (P-ITPA, Esser et al., 2010). This task taps both semantic and lexical skills. The experimenter presents the child with descriptors to elicit progressively more difficult target words, e.g. "I'm thinking about something with feathers, what could that be?" Children are able to score 0, 1 or 2 points for each item and the test is ended early if a child gives a 0-point answer 6 times in a row. Final raw scores were converted to T-scores. Taking the example item "I'm thinking about something that has minutes, what could that be?", a 2-point answer could be "a clock", "time (of day)", "hour", "time (abstract)". These refer to general classifications or superordinates. A 1-point answer could be any specific or concrete example of clocks, time aspects with respect to minutes, or activities and events where measuring time (in minutes) plays an important role, e.g. “a stopwatch” or “alarm clock”.

Productive Grammar

The children’s productive grammar skills were measured with the expressive grammar subtest of the standardised P-ITPA (Esser et al., 2010). In this subtest, the experimenter points to a supporting image and begins a sentence for the child to verbally complete, e.g. “This apple is big, and this is even…[bigger]” or “This is one flower, and these are four… [flowers]”. Children are able to score 0 or 1 points for each item and the test is ended early if a child gives a 0-point answer 6 times in a row. Final raw scores were converted to T-scores. Syntactic and morphological abilities are tested across items, requiring children to build:

- past forms of verbs (simple, perfect, and passive forms; separable verbs are included and correct sentential word order is required);
- comparatives and superlatives of adjectives;
- plurals of nouns with examples of masculine, feminine, and neuter genders, as well as both regular and irregular plurals;
- the genitive form of nouns using possessive determiners;
- case declension of personal pronouns (both accusative and dative forms).

Receptive Grammar

Receptive grammar skills were assessed using the standardised German-language test of grammar comprehension (TROG-D, Fox, 2013). In this test, the experimenter presents the child with a set of 4 images, reads a sentence aloud to the child, and the child must then point to the corresponding target image. This test examines sentence comprehension of a variety of grammatical structures marked by inflection, function words, and sentence structure. These include adjectives, nouns, plurals, verbs, negation, use of the perfect tense and passive voice, topicalization, double object construction, disjunctive conjunctions, and coordination with “and”. Further items cover:

- the prepositions “in”, “on”, “over”, and “under”
- personal pronouns (nominative, accusative and dative)
- 2-element sentences (subject-predicate construction, noun phrase with article and adjective)
- 3-element sentences (subject - predicate - object)
- subordination with "during”, “after”, and “that”
- relative clauses (including pronoun in the accusative or dative cases).

The test contains 84 items, scored in 21 blocks each containing 4 items. All 4 items in the block must be correctly answered to ‘achieve’ a grammatical structure. Their final, total number of correct blocks is then converted into a T-score. The blocks ascend in difficulty, and the test is ended early if at least one out of 4 test items is incorrect across five consecutive blocks. An example test item “The cats look at the ball,” includes 3 distractors alongside the target image: lexical distractor 1 (subject and verb) “The boys are playing with the ball,”; lexical distractor 2 (object) “The cats are looking at the butterfly,”; and grammatical distractor 3 “The cats are looking at the balls.”

Reading

Children’s reading comprehension skills were assessed using a German-language standardised reading screening measure (*Salzburger Lese-Screening für die Schulstufen 2–9* [Salzburg Reading-Screening-Test for School Levels 2–9]: SLS 2–9, Wimmer & Mayringer, 2014). This measure focuses on reading speed and accuracy. It consists of a battery of sentences (ceiling: 100 sentences) that gradually increase in length and complexity. The child must indicate whether each sentence’s content is true or false, e.g. “Trees can speak.” The sentence booklet was posted to each participating child’s home and their caregivers were instructed to hide it until the appointed moment during its online testing session. Before the target test items, the experimenter conducts 2 short practice sessions with 2 separate sets of practice items. The first practice session is not timed and answers are discussed together with the experimenter. The second practice session demonstrates how the timing of the test works. For the scored test, the children are given 3 minutes to complete as many sentences as they can. The items are scored as correct or incorrect and counted. Answers are not scored if they follow 10 mistakes in a row (including not providing an answer). Final raw scores on the test are evaluated relative to a standardised sample of the applicable age. These are then converted to "reading quotients" (RQ) expressing how far the score deviates from the standardisation sample's average. The RQ uses the same scaling as IQ tests but we converted the children's RQs into standard T-scores to compare reading with oral language skills across the descriptive results.

Metaphor Comprehension

We tested the children's comprehension of a total of 22 age-appropriate metaphors embedded within short sentences to provide context (e.g. *Mein Herz ist gebrochen*. [My heart is broken]). A context-embedded approach is in line with prior work (cf. Babarczy et al., 2019; Özçalışkan, 2005). The metaphors were drawn from a larger set originally developed by Vogt and Indefrey (2017) who shared this list with us, permitting its use. Vogt and Indefrey's (2017) set consisted of 11 subcategories of metaphor each containing 5 test items: technomorphic, anthropomorphic, animal, synesthetic (perceptual character), synesthetic (emotional character), pseudo-synesthetic, spatial, process metaphors, action metaphors, action metaphors (emotional states), and condition metaphors (emotional states). To shorten this task to fit our study's time constraints, we conducted a pilot study surveying 30 adult native speakers. We presented the original, full set (11 categories with 5 metaphors each) and asked respondents to identify the two most common metaphors within each category. Using the data from their responses, we removed the extremes (first and least most common) and selected the second most common and second least common metaphors from each category to form our final set of 22 metaphors (c.f. Crawshaw et al., submitted).

For consistency, each metaphor had been pre-recorded by a native speaker and each corresponding audio file was played twice for the children. The experimenter asked the child a sequence of 4 questions: (1) could one say such a sentence; (2) have you heard the sentence before; (3) did you understand it; and if the child answered positively, (4) could you explain the sentence in your own words. If the child answered the third question negatively, then they were asked instead what they thought the sentence might mean. During this task the children were shown a neutral screen consisting of a grey background with a small, black, focusing cross in the centre to hold their attention and minimise unnecessary distraction. The children's answers were recorded, transcribed, and scored as either correct or incorrect on the basis of their explanation of the metaphor in their own words (or what they thought it might mean). A second coder independently scored all responses for 28.6% (a randomly selected 10 of the 35 total participants) of the children for the purpose of calculating interrater reliability (Cohen's $\kappa = .79$).

ToM

The children's theory of mind skills were assessed using two separate measures: (1) the final three subtests of the authorised German-language translation of the Extended Theory-of-Mind Scale (originally by Wellman & Liu, 2004, Peterson et al., 2012; translated by Henning et al., 2013). These covered hidden emotion, sarcasm, and explicit false belief (FB; the unexpected transfer task), all of which were scored as pass or fail. (2) Pre-tested German-language translations of Happé's Strange Stories (following Happé, 1994; and White et al., 2009; pre-tested German translations by Rakoczy et al., 2012, 2018; Ebert, 2020). A total of 8 Strange Stories were evenly distributed across 4 themes: double bluff, white lie, persuasion (deception), and misunderstanding. The target question for each story was either "Why did [they] say this?" or "Why did [they] do this?". Each story was also assigned a neutral (*Kannst du es mir noch ein bisschen genauer sagen?* [Can you say that a bit more precisely for me?]) and mental (e.g. *Was glaubt Simon, was Moritz denkt?* [What does Simon believe Moritz thinks?]) follow-up question. The experimenter used these to clarify the children's responses.

Responses could be scored as correct (2 points), partially correct (1 point), or completely incorrect (0 points), with a potential total of 16 points across the 8 stories. A second coder independently scored all answers for 28.6% of the children (a randomly selected 10 of the 35 total participants) to calculate interrater agreement (ToM-scale: Cohen's κ = 1.0; Strange Stories: Cohen's κ = .92).

The Strange Stories had been pre-recorded by a native speaker and each audio file was played twice for the children. During this task, the children were shown the same grey neutral screen with a fixation cross as in the metaphor comprehension task. To fit the digital context, we updated the stimuli for both the ToM-scale and Strange Stories by making some minor adaptations to the materials. For the ToM-scale, we needed to use the screen-sharing function in the digital conferencing tool to display supporting images which would originally have been physically laid out on the table in front of the child. For the hidden emotion subtest, two separate pictures had to be merged to form one image on the screen: We centred the picture of the story's main character and positioned it above the depictions of the three facial expressions (happy, in-between, and sad). As the context did not support pointing, we adapted the image so that each of the pictured expressions was surrounded by a different and easily distinguishable coloured outline. The children were instead asked to identify the relevant facial expression by the colour of its outline. For the sarcasm subtest, we displayed the original image on the screen but used a spotlight cursor when the experimenter would have pointed to the characters. In the original version of the explicit FB subtest, a small, toy figurine of a boy would be placed on the table near the picture but between the depictions of the story's two locations. For the digital setting, we altered the original image by positioning a cartoon picture of a boy equidistantly between the two location depictions at the bottom of the screen.

The content of two of the eight Strange Stories (German translations: Rakoczy et al., 2012, 2018; Ebert, 2020) were very minimally updated to the modern context but the adaptations did not change the stories' target meanings. One persuasion story was changed so that the protagonist instead tried to obtain pieces of pizza (and not mini sausages). One white-lie story was adapted so that the protagonist's disappointing Christmas present was 'a pile of books' instead of an old brand of encyclopaedia sets that children might not know anymore.

Results

The data was analysed using IBM SPSS Statistics (Version 29) and multiple comparison correction was computed using an online calculator by Hemmerich (2016). We first present the descriptive results of each group across all measures at 9;0 (Table 2). Figure 1 depicts boxplots for the T-score standardised measures across oral language and reading. Table 3 presents the individual performance and language history for each child characterised as a LT at age 2;0. Looking at Figure 1 as well as the groups' mean performances and score ranges in Table 2, we can see overlaps across all of the domains. Group performances appear more comparable across productive grammar, the reading quotient, and the ToM-scale but there are clearer gaps in the domains of productive vocabulary, receptive grammar, Strange Stories, and metaphor comprehension. Many measures were collected during the testing sessions and there are unfortunately a small number of cases with missing data. Sample sizes are therefore reported for each analysis.

Table 2: *Descriptive results of the TD and LT children's mean performance and score range across measures of language, metacognitive, and metalinguistic skills.*

Figure 1: *Results of TD and LT children's performance and score range across the standardised measures of language and reading (T-scores on the y-axis, means, medians, inter quartile ranges, and outliers presented).*

Table 3: *Background and performance of 8 children with a history of LT. a = Size of productive vocabulary measured with FRAKIS (Szagun et al., 2009); b = Standard T-values, measured with SETK-2 (Grimm, 2000); c = taking part in two units with a speech and language pathologist who explained*

and demonstrated language beneficial behaviour; d = did not participate at 5;0; LT = late talker; TD = typical development; DLD = developmental language disorder.

We investigated language (productive vocabulary, productive grammar, receptive grammar), reading, ToM (ToM-scale and Strange Stories), and metaphor comprehension skills in LT versus TD children in our sample. We conducted non-parametric group comparisons due to the uneven and small group sizes. The results presented in Table 4 show that LTs in our sample performed significantly less well than their TD peers on measures of productive vocabulary, receptive grammar, metaphor comprehension, and in the Strange Stories task. The effect sizes of these 4 Mann-Whitney-U tests, measured by the rank-biserial correlation *r* reported in Table 4, were all found to be > 0.4, indicating moderate effects. After applying the Bonferroni-type Benjamini and Hochberg (1995) method to control the false discovery rate when conducting multiple comparisons, we no longer achieved significant group differences on the productive vocabulary measure (Table 4). The LTs trended slightly towards lower performance on the productive grammar measure but this did not reach statistical significance. There were also no statistically significant differences in their performance on the ToM-scale tasks or the reading measure. We ran post hoc power analyses of the Mann-Whitney tests using the software G*Power (Version 3.1.9.4; Faul et al., 2007) for the subtests where the descriptive data indicated lack of power might conceal group differences (productive vocabulary and productive grammar) as well as for subtests showing clearer differences (receptive grammar, the Strange Stories Task, and metaphor comprehension). The following sample sizes would have been required: productive vocabulary (N = 68, TD n = 52, LT n = 16); productive grammar (N = 98, TD n = 76, LT n = 22); receptive grammar (N = 46, TD n = 35, LT n = 11); the Strange Stories task (N = 52, TD n = 41, LT n = 11); and metaphor comprehension (N = 44, TD n = 34, LT n = 10).

Table 4: *Results of the Mann-Whitney-U-tests comparing language, metacognitive and metalinguistic skills between LTs and TD children including performance on each ToM-scale task.*

Bonferroni-type Benjamini and Hochberg (1995) corrections for multiple comparisons and controlling the false discovery rate are reported under p. For the highest level of accuracy, the original reading quotient for the reading score is used instead of our T-score conversion*.

Discussion

There is a paucity of work addressing where 'recovering' LTs might fall between TD and DLD children on potential spectra or continua of language-related competencies. Thus we aimed to address 2 questions within our investigation: (1) Are language and reading abilities impacted in (German-speaking) LTs at age 9;0? (2) Are metacognitive and metalinguistic (ToM and metaphor comprehension) abilities affected in 9;0-year-old children with a history of being a LT? To answer research question (1), our results indicated lower performance on the productive vocabulary and receptive grammar measures but not on the productive grammar and reading measures. To answer research question (2), performance appeared lower on the metaphor comprehension and the advanced ToM Strange Stories tasks but not on the ToM-scale tasks (hidden emotions, sarcasm, and explicit FB). Correcting for multiple comparisons slightly weakened our findings: Group differences in productive vocabulary no longer reached statistical significance. This is likely due to low statistical power resulting from the small number of children with a LT background participating in the study. Despite this limitation, our findings still provide an interesting starting point; studies with larger sample sizes should further investigate ToM and metaphor comprehension alongside language skills in LT children who appear to have "caught up" with their TD peers. We will now address each skill and aim to contemplate (i) why we might have found these differences; and (ii) how our findings might fit within the existing body of research and its relevant theoretical considerations.

Language

We found that LTs seem to catch up with their TD peers in some language skills (productive grammar) but that subtle differences potentially remain across others (vocabulary, receptive grammar). These findings are partially in line with prior work. Rescorla (2002, 2005, 2009) found that LTs performed in the average range on all language tasks at ages 9, 13, and 17 but that their factored scores continued to be significantly lower in vocabulary and grammar, indicating that early delays can have an enduring impact on developmental outcomes. It is not immediately clear why we found no differences in productive grammar but it may be due to the task design of our measures. Rescorla (2009) found that non-language skills measured at age 2 explained some of the variance in her longitudinal sample's performance on vocabulary, grammar, and verbal memory measures. These non-language skills included nonverbal cognitive abilities tapped by tests involving blocks, puzzles, pegs, drawing, and object hiding. They also covered general test-taking skills such as attention, cooperation, and persistence. When these skills were measured at age 17, the LTs scored in the average range but still significantly lower than TD peers (Rescorla, 2009). This could potentially explain our difference with prior work. Our productive grammar task may have tapped different underlying skills than (or may not have been as sensitive as) those used in Rescorla's longitudinal study. Future studies comparing LTs to their TD or DLD peers should carefully consider their task designs and ideally test multiple measures for each language skill to ensure greater sensitivity.

Language within a TD-LT-DLD Spectrum

Given our very small LT sample size, individual differences in this particular group might have meant that they happened to be slightly stronger than other LTs in productive grammar skills, leading to no statistically significant difference with the TD group. This would align with work identifying different subtypes or severities within LTs (Thal et al., 2013) and problems with using LT as a binary clinical category (Dollaghan, 2013). It would also fit with prior theories (Bates et al., 1995; Dollaghan, 2013; Rescorla, 2009, 2011; Thal et al., 2013) regarding a dimensional account of language delay where individual LT and TD children differ quantitatively along a language ability spectrum. Considering the post hoc power analyses, it is however also possible that the group differences in productive grammar might be so small that our sample size was insufficient to identify it.

Reading

We found no significant difference in reading comprehension between LT and TD groups at age 9. This contrasts with some prior work: LTs in Rescorla's (2002, 2005) longitudinal sample scored significantly lower on reading comprehension at age 13 but showed no difference at ages 6 or 7 and only began to show significant differences when they were tested at ages 8 and 9. Differences in task design and demands could be a factor here but another point also worth considering is cross-linguistic difference in orthographic transparency. English is an orthographically less transparent language than German (the language of the participants in our study). It is possible that differences in reading comprehension performance may become evident earlier in English readers than German readers (Borleffs et al., 2019; Diamanti et al., 2018). Prior cross-linguistic work with English- and German-speaking dyslexic children aged 10–12 found that the English-speaking group was more greatly impaired in reading accuracy than the German-speaking one (Landerl et al., 1997). This might explain why we found no significant difference in our sample's performance at age 9. Our participants might not have yet reached a developmental point where underlying differences in language(-adjacent) skills had begun to accelerate disparities in reading comprehension outcomes. Alternatively, the task design might need to be more complex and possess enhanced sensitivity to identify any differences between German-speaking TD and LT children. These possibilities should definitely be investigated further in future work with longitudinal samples of German-speaking LTs.

Dyslexia

Some prior work (Psyridou et al., 2018) has identified that a family history of dyslexia is a stronger predictor than LT status for reading comprehension performance. We would like to note here that our analyses could have been based on partial information. In our sample of 34 children, 1 of our 8 LTs and 3 of our 26 TD children had been diagnosed with dyslexia. However, 6 (only 1 of whom was a LT) of the 30 non-dyslexic children performed more poorly than the highest performing dyslexic child. The other 6 non-dyslexic LTs' reading scores were clustered at or above the median. Some of the children might possibly have had undiagnosed dyslexia and might go on to receive a diagnosis in future. Once the study finished, all of the participating children's parents were informed about their child's comparative performances across the tasks. They were given advice for how to support their child's development in these areas so that they could seek further assistance or assessment.

Reading within a TD-LT-DLD Spectrum

Psyridou et al. (2018) noted that only LTs who had experienced both expressive and receptive language delay (late comprehender group in Thal et al., 2013) went on to face difficulties with reading comprehension, and not LTs who had experienced expressive without receptive language delay (late producer group in Thal et al., 2013). Bishop, Edmundson, Stothard and colleagues' (1987, 1998) longitudinal sample performed similarly. The 'good-outcome' LT group (language difficulties appeared to resolve) had a lower median performance than the controls on verbal comprehension and reading, while the 'poor-outcome' LT group had an even lower median performance (Stothard et al., 1998). This could support the idea that one subgroup of LTs is more likely to experience challenges in meaning-based decoding (Thal et al., 2013). Unfortunately, this potential split cannot be addressed within our study due to the small LT sample size, however future work should account for these subgroupings when considering measures that tap meaning-based decoding.

Advanced Language-Related Skills: ToM and Metaphor Comprehension

We found that LTs appeared to perform less well than their TD peers on the Strange Stories and metaphor comprehension tasks but not on the ToM-scale (hidden emotions, sarcasm, and explicit FB). The three ToM-scale subtests could only be scored as pass or fail so it is difficult to build up a more nuanced picture of the abilities tested and their age-appropriateness. The explicit FB and hidden emotions tasks were actually designed so that much younger children (around 4-5 years of age) could pass them. Our main reason for including them in our test battery was to conduct a baseline assessment of children's abilities. Prior research (c.f. Peterson et al., 2012) has also shown that sarcasm is still emerging at age 9, so we may have assessed it slightly too early in our sample. Testing it a little later on when it is more established might better reflect each group's abilities.

ToM and Metaphor Comprehension within a TD-LT-DLD Spectrum

To our knowledge, no other studies have investigated ToM and metaphor comprehension abilities in LTs who appear to 'recover'. We can therefore only attempt to fit our findings somewhere between and within the existing work on TD and DLD children. There are different possibilities to explore in seeking an explanation for the LTs' comparatively lower performance on the Strange Stories and metaphor comprehension tasks. One starting point could be the LTs' lower performance on vocabulary and receptive grammar measures. Prior work has associated these abilities with performance on ToM and metaphor comprehension tasks (see e.g. Deckert et al., 2019; Milligan et al., 2007; Osterhaus & Koerber, 2021). Combining the ideas of a language endowment spectrum (Ellis Weismer, 2007) and language ability as a dimensional spectrum along which individuals differ (Rescorla, 2009, 2011), we could further incorporate language-associated ToM and metaphor comprehension skills. It would be logical for LTs without a DLD diagnosis to fall somewhere along the spectrum (or continuum) between their TD and DLD peers in these competencies too.

Potential Differences in ToM Task Demands

LTs who do go on to receive a DLD diagnosis have been found to struggle with language processing (Sansavini et al., 2021). It is therefore also possible that our audio-based Strange Stories and metaphor comprehension tasks may have been harder for the LTs to process than the ToM-scale tasks, which were supported visually. However, Nilsson and de Lopez (2016) have reviewed numerous studies in which children with DLD showed ToM impairments even when completing tasks with lower verbal demands. This suggests that at least for DLD, verbal processing demands cannot necessarily explain away ToM impairments. Nevertheless, rather than the 'social-perceptual' component of ToM (Tager-Flusberg and Joseph, 2005), Nilsson and de Lopez (2019) proposed that children with DLD's ToM delays are rooted in the 'socio-cognitive' ToM component (Tager-Flusberg & Joseph, 2005). This component supposedly enables inference-making about mental phenomena through the integration of different kinds of information. This could tie into both a processing-deficit perspective on LTs and children with DLD as well as the idea of a language and language-related ability spectrum or continuum. Relevant for DLD-adjacent LTs and irrespective of a ToM deficit versus delay, Nilsson and de Lopez (2019) importantly highlight that potential impairments early on could become enduring characteristics that affect continued social development and outcomes.

Metaphor Comprehension and Processing

Work has shown that metaphor comprehension is important for peer social interaction outcomes (Del Sette et al., 2021) and that it is linked to both language (e.g. Deckert et al., 2019) and ToM ability (Norbury, 2005). Understanding ambiguous metaphors can require thinking about a speaker's intended meaning, and some metaphors in themselves contain mental aspects (Lecce et al., 2019). LTs' lower performance on the metaphor comprehension task would be consistent with prior work investigating metaphor comprehension in DLD populations (Bühler et al., 2018; Norbury, 2005; Spanoudis, 2016). While task demands in terms of verbal processing could have been a contributing factor, the findings from Lorusso et al.'s (2015) study using Event-Related Potentials (ERP) could also provide some interesting insights. Lorusso et al. (2015) compared children with DLD, children with nonverbal learning disabilities, and a control TD group (aged 6–15, matched for both gender and chronological age) on the processing of literal versus figurative uses of verbs (e.g. *I picked up a flower* vs. *I picked up an idea*). This is analogous to metaphor comprehension. They found that the children with DLD showed the lowest accuracy and the most deviant ERP patterns when processing abstract and figurative expressions (Lorusso et al., 2015). They also performed worse on abstract sentences than the participants with nonverbal learning disabilities (Lorusso et al., 2015). The authors argued that this lent "…support to theories assuming that linguistic processes are crucial to the representation and comprehension of abstract language" (Lorusso et al., 2015, p.17). These results are meaningful for our findings regarding LTs, reinforcing the idea that underlying impairments in language ability may critically disrupt non-literal language processing.

Looking to the Future

As children grow older, metaphor comprehension and ToM skills become increasingly important for academic and social success. Our findings indicate that LTs may need additional support with these competencies to ensure they do not fall behind their TD peers at school. However, despite these potential deficits, the highest scorer on the metaphor comprehension task in our study was actually a LT (late producer but not comprehender; see tables 2 and 3). This highlights that rather than a binary LT status, more nuance is required for advanced metalinguistic skills. To develop appropriate, tailored support for LTs, future work needs to go beyond basic differences in performance outcomes. It must address remaining questions about whether (a) metalinguistic and metacognitive competencies are impacted simply as a result of impaired language skills; or (b) language and these competencies share common underlying deficits; or even (c) there are associated but differing impairments across both these competencies and language ability.

Limitations

Despite our longitudinal perspective and longstanding interaction with the participating children which strengthen this study's findings, we must address some limitations. Firstly, we recognise that this study is based on a small sample size (N = 35) and that the participating groups of TD and LT children are uneven (N = 35; TD n = 27, LT n = 8) with a very small number of participating LTs. A definitive generalisation from our results is consequently not possible. Instead they provide initial insights into whether LTs perform differently compared to their TD peers on metalinguistic (metaphor comprehension) and advanced metacognitive (Strange Stories) tasks. More conclusive work is needed with matched, larger, and even groups of participants. Each target ability should be tested using a variety of measures to ensure that contrasting findings do not arise from task differences between studies.

Secondly, although all previous testing points in our wider longitudinal study were carried out in-person, we had to adapt in response to the COVID-19 pandemic and implement a hybrid testing method. Two sessions were carried out digitally and one session in person. Prior to the pandemic, it was fairly uncommon to use digital testing rather than the standard in-person experimental method. This could perhaps weaken our results. There are challenges involved in using online testing methods (Braun et al., 2020) as well as tools for research such as Zoom (Gray et al., 2020). However, there are also early indications that studies carried out via Zoom with young children (Escudero et al., 2021) can still be an appropriate, reliably comparable, alternative measure to testing in-person. Of course, further evaluation is still needed. Some additional, possibly mitigating factors include the highly positive reception of our hybrid method by the participating children and families, as well as the longstanding in-person context and relationship built up with them by the university department. This context afforded us a well-established developmental picture of the children's abilities, lending weight to the potential reliability of our findings.

Implications and Future Opportunities

Our work indicates that there may be a spectrum and even a complex continuum of language ability and associated skills, such as metalinguistic and metacognitive competencies. LTs can move past deficits in performance which would lead to a classification of "disorder" and appear to catch up, yet still demonstrate discernible differences from TD children many years later. Metalinguistic and metacognitive competencies are important for effective social interaction with peers and enable engagement with more complex academic tasks like the study of literature. Our findings suggest that even LTs who appear to 'recover' might benefit from targeted, extra support. This could be provided informally by caregivers or more formally within intervention programmes or in a school context. Future comparative research should investigate these skills within larger sized samples, integrating children's history of language development and building (sub-)groups of TD, LT, and DLD children. More finely grained approaches are needed to identify which aspects of advanced metacognitive and metalinguistic abilities are specifically impacted in these groups, along with their associated linguistic competencies and cognitive functions.

Acknowledgments

We thank the German Research Foundation (DFG) for funding the earlier stages of this longitudinal study between the years 2012–2018 under the grant numbers (RI 898/6-1 and RI 898/6-2 to UR and LU 2116/1-2 to CL) which was then extended by internal money from the authors' university department. We are very grateful to all the families who have continued to participate over these many years as well as to our collaborators Angela Grimminger, Katharina Rohlfing, and Ulf Liszkowski who were involved in the project during previous stages. We are also very thankful for all the research assistants who over the years have assisted with or have been involved in carrying out our longitudinal study, the preparation of materials, or in the transcription and coding process: Jannika Böse, Annika Büssing, Olivia Espeter, Tabea Musshoff, Katharina Rademacher, Kathrin Gremplewski, Lisa Grundhöfer, Inga Hackbarth, Bianca Henze, Janina Jurjahn, Miriam Krimer, Carina Lucas, Laura Polke, Anika Queißer, Suzan Ryschka, and Annika Biewener.

Data Availability Statement

The dataset generated for this study cannot be made publicly available as participants did not provide consent for that, however private enquiries to view an anonymised version of the dataset are welcomed and can be addressed to the corresponding author.

References

Andrés-Roqueta, C., Adrian, J. E., Clemente, R. A., & Katsos, N. (2013). Which are the best predictors of theory of mind delay in children with specific language impairment?. *International Journal of*

Language & Communication Disorders, *48*(6), 726–737. https://doi.org/10.1111/1460-6984.12045

Astington, J. & Baird, J. (2005). Introduction: Why Language Matters. In Astington, J., & Baird, J. (Eds.), *Why Language Matters for Theory of Mind* (pp.3–25). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195159912.003.0001

Atkinson, L., Slade, L., Powell, D., & Levy, J. P. (2017). Theory of mind in emerging reading comprehension: A longitudinal study of early indirect and direct effects. *Journal of Experimental Child Psychology*, *164*, 225–238. https://doi.org/10.1016/j.jecp.2017.04.007

Babarczy, A., Balázs, A., & Krizsai, F. (2019). Preschoolers' Metaphor Comprehension. Methodological Issues in Experimental Pragmatics. *Acta Universitatis Sapientiae, Philologica*, *11*(2), 133–150. https://doi.org/10.2478/ausp-2019-0017

Bates, E., Dale, P., & Thal, D. (1995). Individual differences and their implications for theories of language development. In P. Fletcher & B. MacWhinney (Eds.), *Handbook of Child Language* (pp. 96–151). Oxford: Basil Blackwell. https://doi.org/10.1111/b.9780631203124.1996.00005.x

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological), 57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bishop, D. V. M. (2017). Why is it so hard to reach agreement on terminology? The case of developmental language disorder (DLD). *International Journal of Language & Communication Disorders, 52*(6), 671–680. https://doi.org/10.1111/1460-6984.12335

Bishop, D. V. M., & Edmundson, A. (1987). Language-impaired 4-year-olds: Distinguishing transient from persistent impairment. *Journal of Speech and Hearing Disorders, 52*(2), 156–173. https://doi.org/10.1044/jshd.5202.156

Bishop, D. V., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & Catalise-2 Consortium. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry, 58*(10), 1068–1080. https://doi.org/10.1111/jcpp.12721

Boerma, I. E., Mol, S. E., & Jolles, J. (2017). The role of home literacy environment, mentalizing, expressive verbal ability, and print exposure in third and fourth graders' reading comprehension. *Scientific Studies of Reading, 21*, 179–193. https://doi.org/10.1080/10888438.2016.1277727

Borleffs, E., Maassen, B. A., Lyytinen, H., & Zwarts, F. (2019). Cracking the code: The impact of orthographic transparency and morphological-syllabic complexity on reading and developmental dyslexia. *Frontiers in Psychology, 9*, 2534. https://doi.org/10.3389/fpsyg.2018.02534

Braun, R., Blok, V., Loeber, A., & Wunderle, U. (2020). COVID-19 and the onlineification of research: kick-starting a dialogue on Responsible online Research and Innovation (RoRI). *Journal of Responsible Innovation*, *7*(3), 680–688. https://doi.org/10.1080/23299460.2020.1789387

Bühler, D., Perovic, A., & Pouscoulous, N. (2018). Comprehension of novel metaphor in young children with Developmental Language Disorder. *Autism & Developmental Language Impairments*, *3*, 1–11. https://doi.org/10.1177/2396941518817229

Crawshaw, C. E., Lüke, C., & Ritterfeld, U. (2024). *Does Early Gesture Usage Contribute Alongside Oral Language to Later Theory of Mind Performance and Metaphor Comprehension? Indications from a Longitudinal Study with Children Aged 1–9 Years.* Manuscript submitted for publication.

Deckert, M., Schmoeger, M., Schaunig-Busch, I., & Willinger, U. (2019). Metaphor processing in middle childhood and at the transition to early adolescence: The role of chronological age, mental age, and verbal intelligence. *Journal of Child Language*, *46*(2), 334–367. https://doi.org/10.1017/S0305000918000491

Del Sette, P., Ronchi, L., Bambini, V., & Lecce, S. (2021). Longitudinal associations between metaphor understanding and peer relationships in middle childhood. *Infant and Child Development*, e2232. https://doi.org/10.1002/icd.2232

De Mulder, H. N. M., Wijnen, F., & Coopmans, P. H. A. (2019). Interrelationships between Theory of Mind and language development: A longitudinal study of Dutch-speaking kindergartners. *Cognitive Development*, *51*, 67–82. https://doi.org/10.1016/j.cogdev.2019.03.006

De Villiers, J. (2007). The interface of language and theory of mind. *Lingua*, *117*(11), 1858–1878. https://doi.org/10.1016/j.lingua.2006.11.006

De Villiers, P. A. (2005). The Role of Language in Theory-of-Mind Development: What Deaf Children Tell Us. In Astington, J., & Baird, J. (Eds.), *Why Language Matters for Theory of Mind* (pp. 266–297). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195159912.003.0013

Devine, R. T., & Lecce, S. (Eds.). (2021). *Theory of Mind in Middle Childhood and Adolescence: Integrating Multiple Perspectives*. Routledge. https://doi.org/10.4324/9780429326899

Devine, R. T., White, N., Ensor, R., & Hughes, C. (2016). Theory of mind in middle childhood: Longitudinal associations with executive function and social competence. *Developmental Psychology*, *52*(5), 758–771. https://doi.org/10.1037/dev0000105

Diamanti, V., Goulandris, N., Campbell, R., & Protopapas, A. (2018). Dyslexia profiles across orthographies differing in transparency: An evaluation of theoretical predictions contrasting English and Greek. *Scientific Studies of Reading, 22*(1), 55–69. https://doi.org/10.1080/10888438.2017.1338291

Doherty, M. (2008). *Theory of mind: How children understand others' thoughts and feelings*. Psychology Press. https://doi.org/10.4324/9780203929902

Dollaghan, C. (2013). Late talker as a clinical category: A critical evaluation. In L. A. Rescorla & P. S. Dale (Eds.), *Late talkers: Language development, interventions, and outcomes* (pp. 91–112). Baltimore, MD: Brookes.

Durrleman, S., Burnel, M., & Reboul, A. (2017). Theory of mind in SLI revisited: links with syntax, comparisons with ASD. *International Journal of Language & Communication Disorders*, *52*(6), 816–830. https://doi.org/10.1111/1460-6984.12317

Durrleman, S., & Delage, H. (2020). Training Complements for Belief Reasoning in Developmental Language Disorder. *Journal of Speech, Language, and Hearing Research*, *63*(6), 1861–1877. https://doi.org/10.1044/2020_JSLHR-19-00075

Ebert, S. (2020). Theory of mind, language, and reading: Developmental relations from early childhood to early adolescence. *Journal of Experimental Child Psychology*, *191*, 104739. https://doi.org/10.1016/j.jecp.2019.104739

Ensor, R., Devine, R. T., Marks, A., & Hughes, C. (2014). Mothers' cognitive references to 2-year-olds predict theory of mind at ages 6 and 10. *Child development*, *85*(3), 1222–1235. https://doi.org/10.1111/cdev.12186

Ellis Weismer, S. (2007). Typical talkers, late talkers, and children with specific language impairment: A language endowment spectrum? In R. Paul (Ed.), *The influence of developmental perspectives on research and practice in communication disorders: A festschrift for Robin S. Chapman* (pp. 83–102). Erlbaum.

Escudero, P., Pino Escobar, G., Casey, C. G., & Sommer, K. (2021). Four-Year-Old's Online Versus Face-to-Face Word Learning via eBooks. *Frontiers in Psychology*, *12*, 450. https://doi.org/10.3389/fpsyg.2021.610975

Esser, G., Wyschkon, A., Ballaschk, K., & Hänsch, S. (2010). *Potsdam-Illinois Test für Psycholinguistische Fähigkeiten – P-ITPA* [Potsdam-Illinois Test for Psycholinguistic Abilities – P-ITPA]. Hogrefe, Testzentrale.

Farrant, B. (2015). Specific language impairment and perspective taking: Delayed development of theory of mind, visual and emotional perspective taking. *Journal of Childhood & Developmental Disorders*, *1*(1), 1–8. https://doi.org/10.4172/2472-1786.100008

Farrant, B. M., Fletcher, J., & Maybery, M. T. (2006). Specific language impairment, theory of mind, and visual perspective taking: Evidence for simulation theory and the developmental role of language. *Child Development*, *77*(6), 1842–1853. https://doi.org/10.1111/j.1467-8624.2006.00977.x

Farrar, M. J., Johnson, B., Tompkins, V., Easters, M., Zilisi-Medus, A., & Benigno, J. P. (2009). Language and theory of mind in preschool children with specific language impairment. *Journal of Communication Disorders*, *42*(6), 428-441. https://doi.org/10.1016/j.jcomdis.2009.07.001

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/bf03193146

Fox, A. V. (2013). *TROG-D. Test zur Überprüfung des Grammatikverständnisses* [Test for Assessing Grammatical Understanding] 6th Edition. Schulz-Kirchner.

Gillott, A., Furniss, F., & Walter, A. (2004). Theory of mind ability in children with specific language impairment. *Child Language Teaching and Therapy*, *20*(1), 1–11. https://doi.org/10.1191/0265659004ct260oa

Gordon Pershey, M. (2000). Children's Elicited Use of Pragmatic Language Functions: How Six-and Seven-year-old Children Adapt to the Interactional Environments of Story Scenarios. *Language Awareness, 9*(4), 218–235. https://doi.org/10.1080/09658410008667147

Gray, L. M., Wong-Wylie, G., Rempel, G. R., & Cook, K. (2020). Expanding qualitative research interviewing strategies: Zoom video communications. *The Qualitative Report*, *25*(5), 1292–1301. https://doi.org/10.46743/2160-3715/2020.4212

Grimm, H. (2000). *SETK-2. Sprachentwicklungstest für zweijährige Kinder.* Göttingen: Hogrefe.

Grimm, H. (2001). *Sprachentwicklungstest für drei- bis fünfjährige Kinder (SETK 3-5): Diagnose von Sprachverarbeitungsfähigkeiten und auditiven Gedächtnisleistungen.* Göttingen: Hogrefe.

Hammer, C. S., Morgan, P., Farkas, G., Hillemeier, M., Bitetti, D., & Maczuga, S. (2017). Late talkers: A population-based study of risk factors and school readiness consequences. *Journal of Speech, Language, and Hearing Research, 60*(3), 607–626. https://doi.org/10.1044%2F2016_JSLHR-L-15-0417

Happé, F. G. E. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders*, *24*(2), 129–154. https://doi.org/10.1007/BF02172093

Happé, F. G. E. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development*, *66*(3), 843–855. https://doi.org/10.1111/j.1467-

8624.1995.tb00909.x

Hemmerich, W. (2016). StatistikGuru: Rechner zur Adjustierung des α-Niveaus [*StatisticGuru: Calculator for adjusting the α level*]. https://statistikguru.de/rechner/adjustierung-des-alphaniveaus.html

Henning, A., Hofer, T., & Aschersleben, G. (2013). *Erweiterte "Theory of Mind-Skala" für 3- bis 11-jährige Kinder* [Extended Theory-of-Mind Scale for 3- to 11-year-old children], 2nd Edition. Universität des Saarlandes Arbeitseinheit Entwicklungspsychologie.

Hjetland, H. N., Lervåg, A., Lyster, S. A. H., Hagtvet, B. E., Hulme, C., & Melby-Lervåg, M. (2019). Pathways to reading comprehension: A longitudinal study from 4 to 9 years of age. *Journal of Educational Psychology*, *111*(5), 751–763. http://dx.doi.org/10.1037/edu0000321

Horvath, S., Kueser, J. B., Kelly, J., & Borovsky, A. (2022). Difference or delay? Syntax, semantics, and verb vocabulary development in typically developing and late-talking toddlers. *Language Learning and Development, 18*(3), 352–376. https://doi.org/10.1080/15475441.2021.1977645

Kalandadze, T., Norbury, C., Nærland, T., & Næss, K. A. B. (2018). Figurative language comprehension in individuals with autism spectrum disorder: A meta-analytic review. *Autism, 22*(2), 99–117. https://doi.org/10.1177/1362361316668652

Kauschke, C., and Siegmüller, J. (2010). *Patholinguistische Diagnostik bei Sprachentwicklungsstörungen (PDSS)*, 2nd Edn. München: Urban & Fischer.

Kirk, E., Donnelly, S., Furman, R., Warmington, M., Glanville, J., & Eggleston, A. (2022). The relationship between infant pointing and language development: A meta-analytic review. *Developmental Review, 64*, 101023. https://doi.org/10.1016/j.dr.2022.101023

Landerl, K., Wimmer, H., & Frith, U. (1997). The impact of orthographic consistency on dyslexia: A German-English comparison. *Cognition, 63*(3), 315–334. https://doi.org/10.1016/S0010-0277(97)00005-X

Language and Reading Research Consortium, & Chiu, Y. D. (2018). The simple view of reading across development: Prediction of grade 3 reading comprehension from prekindergarten skills.

Remedial and Special Education, *39*(5), 289–303. https://doi.org/10.1177/0741932518762055

Lecce, S., Ronchi, L., Del Sette, P., Bischetti, L., & Bambini, V. (2019). Interpreting physical and mental metaphors: Is Theory of Mind associated with pragmatics in middle childhood? *Journal of Child Language, 46*(2), 393–407. https://doi.org/10.1017/S030500091800048X

Levorato, M. C., Nesi, B., & Cacciari, C. (2004). Reading comprehension and understanding idiomatic expressions: A developmental study. *Brain and Language*, *91*(3), 303–314. https://doi.org/10.1016/j.bandl.2004.04.002

Lorusso, M. L., Burigo, M., Borsa, V., & Molteni, M. (2015). Processing sentences with literal versus figurative use of verbs: an ERP study with children with language impairments, nonverbal impairments, and typical development. *Behavioural Neurology*, *2015,* 475271, 1–21. https://doi.org/10.1155/2015/475271

Lüke, C., Grimminger, A., Rohlfing, K. J., Liszkowski, U., & Ritterfeld, U. (2017). In infants' hands: Identification of preverbal infants at risk for primary language delay. *Child Development, 88*(2), 484–492. https://doi.org/10.1111/cdev.12610

Lüke, C., Kauschke, C., Dohmen, A., Haid, A., Leitinger, C., Männel, C., ... & Neumann, K. (2023). Definition and terminology of developmental language disorders—Interdisciplinary consensus across German-speaking countries. *PLoS ONE, 18*(11), e0293736. https://doi.org/10.1371/journal.pone.0293736

Lüke, C., Ritterfeld, U., Grimminger, A., Rohlfing, K. J., & Liszkowski, U. (2020). Integrated Communication System: Gesture and Language Acquisition in Typically Developing Children and Children with LD and DLD. *Frontiers in Psychology*, *11*, 118. https://doi.org/10.3389/fpsyg.2020.00118

Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Development*, *78*(2), 622–646. https://doi.org/10.1111/j.1467-8624.2007.01018.x

Nilsson, K. K., & de Lopez, K. J. (2016). Theory of mind in children with specific language impairment:

A systematic review and meta-analysis. *Child Development*, *87*(1), 143–153. https://doi.org/10.1111/cdev.12462

Norbury, C. F. (2005). The relationship between theory of mind and metaphor: Evidence from children with language impairment and autistic spectrum disorder. *British Journal of Developmental Psychology*, *23*(3), 383–399. https://doi.org/10.1348/026151005X26732

Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vamvakas, G., & Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: Evidence from a population study. *Journal of Child Psychology and Psychiatry, 57*(11), 1247–1257. https://doi.org/10.1111/jcpp.12573

Osterhaus, C., & Koerber, S. (2021). The development of advanced theory of mind in middle childhood: A longitudinal study from age 5 to 10 years. *Child Development, 92*(5), 1872–1888. https://doi.org/10.1111/cdev.13627

Özçalışkan, Ş. (2005). On learning to draw the distinction between physical and metaphorical motion: Is metaphor an early emerging cognitive and linguistic capacity? *Journal of Child Language*, *32*(2), 291. https://doi.org/10.1017/S0305000905006884

Paul, R., & Ellis Weismer, S. (2013). Late talking in context: The clinical implications of delayed language development. In L. Rescorla & P. S. Dale (Eds.), *Late talkers: Language development, interventions, and outcomes* (pp. 203–217). Baltimore, MD: Brookes.

Perry, L. K., Kucker, S. C., Horst, J. S., & Samuelson, L. K. (2023). Later bloomer or language disorder? Differences in toddler vocabulary composition associated with long-term language outcomes. *Developmental Science, 26*(4), e13342. https://doi.org/10.1111/desc.13342

Peterson, C. C., Wellman, H. M., & Slaughter, V. (2012). The mind behind the message: Advancing theory-of-mind scales for typically developing children, and those with deafness, autism, or Asperger syndrome. *Child Development*, *83*(2), 469–485. https://doi.org/10.1111/j.1467-8624.2011.01728.x

Pronina, M., Prieto, P., Bischetti, L., & Bambini, V. (2023). Expressive pragmatics and prosody in young preschoolers are more closely related to structural language than to mentalizing. *Language Learning and Development, 19*(3), 323–344. https://doi.org/10.1080/15475441.2022.2074852

Psyridou, M., Eklund, K., Poikkeus, A. M., & Torppa, M. (2018). Reading outcomes of children with delayed early vocabulary: A follow-up from age 2–16. *Research in Developmental Disabilities*, *78*, 114–124. https://doi.org/10.1016/j.ridd.2018.05.004

Rakhlin, N., Kornilov, S. A., Reich, J., Babyonyshev, M., Koposov, R. A., & Grigorenko, E. L. (2011). The relationship between syntactic development and theory of mind: Evidence from a small-population study of a developmental language disorder. *Journal of Neurolinguistics*, *24*(4), 476–496. https://doi.org/10.1016/j.jneuroling.2011.03.001

Rakoczy, H., Wandt, R., Thomas, S., Nowak, J., & Kunzmann, U. (2018). Theory of mind and wisdom: The development of different forms of perspective-taking in late adulthood. *British Journal of Psychology*, *109*(1), 6–24. https://doi.org/10.1111/bjop.12246

Rakoczy, H., Harder‐Kasten, A., & Sturm, L. (2012). The decline of theory of mind in old age is (partly) mediated by developmental changes in domain-general abilities. *British Journal of Psychology*, *103*(1), 58–72. https://doi.org/10.1111/j.2044-8295.2011.02040.x

Rescorla, L. A. (2013). Late-talking toddlers: A 15-year follow-up. In L. A. Rescorla & P. S. Dale (Eds.), *Late talkers: Language development, interventions, and outcomes* (pp. 219–239). Baltimore, MD: Brookes.

Rescorla, L. (2009). Age 17 language and reading outcomes in late-talking toddlers: Support for a dimensional perspective on language delay. *Journal of Speech, Language, and Hearing Research, 52*, 16–30. https://doi.org/10.1044/1092-4388(2008/07-0171)

Rescorla, L. (2005). Age 13 language and reading outcomes in late-talking toddlers. *Journal of Speech, Language, and Hearing Research, 48*, 459–472. https://doi.org/10.1044/1092-4388(2005/031)

Rescorla, L. (2002). Language and reading outcomes to age 9 in late-talking toddlers. *Journal of*

Speech, Language, and Hearing Research, 45, 360–371. https://doi.org/10.1044/1092-4388(2002/028)

Rescorla, L., Roberts, J., & Dahlsgaard, K. (1997). Late talkers at 2: Outcome at age 3. *Journal of Speech, Language, and Hearing Research, 40*, 556–566. https://doi.org/10.1044/jslhr.4003.556

Sansavini, A., Favilla, M. E., Guasti, M. T., Marini, A., Millepiedi, S., Di Martino, M. V., ... & Lorusso, M. L. (2021). Developmental Language Disorder: Early Predictors, Age for the Diagnosis, and Diagnostic Tools. A Scoping Review. *Brain Sciences*, *11*(5), 654–691. https://doi.org/10.3390/brainsci11050654

Smit, L., Knoors, H., Hermans, D., Verhoeven, L., & Vissers, C. (2019). The interplay between theory of mind and social emotional functioning in adolescents with communication and language problems. *Frontiers in Psychology*, *10*, 1488. https://doi.org/10.3389/fpsyg.2019.01488

Spanoudis, G. (2016). Theory of mind and specific language impairment in school-age children. *Journal of Communication Disorders*, *61*, 83–96. https://doi.org/10.1016/j.jcomdis.2016.04.003

Stothard, S. E., Snowling, M. J., Bishop, D. V. M., Chipchase, B. B., & Kaplan, C. A. (1998). Language-impaired preschoolers: A follow-up into adolescence. *Journal of Speech, Language, and Hearing Research, 41*(2), 407–418. https://doi.org/10.1044/jslhr.4102.407

Szagun, G., Stumper, B., and Schramm, S. A. (2009). *FRAKIS. Fragebogen zur frühkindlichen Sprachentwicklung*. Frankfurt am Main: Pearson.

Tager-Flusberg, H., & Joseph, R. M. (2005). How language facilitates the acquisition of false belief understanding in children with autism. In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 298–319). New York, NY: Oxford University Press. https://psycnet.apa.org/doi/10.1093/acprof:oso/9780195159912.003.0014

Tellegen, P. J., Laros, J. A., and Petermann, F. (2007). *SON-R 2 1/2–7: Non-Verbaler Intelligenztest [Non-verbal intelligence test]*. Hogrefe.

Thal, D. J., Marchman, V. A., & Tomblin, J. B. (2013). Late-talking toddlers: Characterization and prediction of continued delay. In L. A. Rescorla & P. S. Dale (Eds.), *Late talkers: Language development, interventions, and outcomes* (pp. 169–201). Baltimore, MD: Brookes.

Vissers, C., & Koolen, S. (2016). Theory of mind deficits and social emotional functioning in preschoolers with specific language impairment. *Frontiers in Psychology, 7*, 1–7. https://doi.org/10.3389/fpsyg.2016.01734

Vogt, S., & Indefrey, P. (2017). Metaphernerwerb: eine empirische Studie bei Kindern im Alter von sechs bis vierzehn Jahren [Metaphor acquisition: an empirical study with children between the ages of six to fourteen years]. *Metaphorik.de*, 27/2017, 69–106. https://www.metaphorik.de/de/journal/27/metaphernerwerb-eine-empirische-studie-bei-kindern-im-alter-von-sechs-bis-vierzehn-jahren.html

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, *75*(2), 523–541. https://doi.org/10.1111/j.1467-8624.2004.00691.x

White, S., Hill, E., Happé, F., & Frith, U. (2009). Revisiting the strange stories: Revealing mentalizing impairments in autism. *Child Development*, *80*(4), 1097–1117. https://doi.org/10.1111/j.1467-8624.2009.01319.x

Wimmer, H., & Mayringer, H. (2014). *Salzburger Lese-Screening für die Schulstufen 2–9: SLS 2–9* [Salzburg Reading-Screening-Test for School Levels 2–9: SLS 2–9]. Huber.

Zubrick, S. R., Taylor, C. L., Rice, M. L., & Slegers, D. W. (2007). Late language emergence at 24 months: An epidemiological study of prevalence, predictors, and covariates. *Journal of Speech, Language, and Hearing Research, 50*, 1562–1592. https://doi.org/10.1044/1092-4388(2007/106)

Table 1: Overview of recent studies considering ToM performance in children with DLD.

Authors	Participants	ToM Task(s) Used	Findings re. ToM
Andrés-Roqueta et al., 2013	Children with DLD (*n* = 31) aged 3;5–7;5, age- and gender-matched TD controls (*n* = 31), younger language- and gender-matched TD controls (*n* = 31)	2 False belief (FB) tasks: unexpected transfer and unexpected contents	DLD group performed worse than age-matched group but similarly to language-matched group
Durrleman et al., 2017	Children with DLD (*n* = 20) aged (6;5–11;7), children on autistic spectrum (*n* = 34) aged 6;9–14;4), TD children (*n* = 30) aged 4;9–11;8	Low-verbal picture-sequencing task with a story requiring FB attribution	Children across groups performed similarly depending on their sentential complementation ability
Farrant, 2015	Children with DLD (*n* = 30) aged 4;0–6;2, TD children (*n* = 30) matched for non-verbal ability, gender, and age	Diverse desires, diverse beliefs, knowledge access, FB: unexpected contents, low verbal FB, visual perspective-taking (VPT), emotional perspective-taking	DLD group performed worse than TD group on all measures except for diverse beliefs (difference marginally significant)
Farrant et al., 2006	Children with DLD (*n* = 20) aged 4;10–5;8, TD children (*n* = 20) matched for non-verbal ability, gender, and age	Diverse desires, diverse beliefs, knowledge access, FB: unexpected contents, real–apparent emotions, VPT	DLD group performed worse than TD group on more complex VPT, knowledge access, and FB: unexpected contents
Farrar et al., 2009	Children with DLD (*n* = 34) aged 3;6–5;5	FB: unexpected contents, FB: unexpected transfer, 3 appearance–reality: 2 mistaken attribute, 1 mistaken identity	General grammatical development and vocabulary predicted ToM ability; sentential complementation no unique role
Gillott et al., 2004	Children with DLD (*n* = 15), children with high-functioning autism (*n* = 15), and TD children (*n* = 15): aged 8–12, age- and gender matched	12 Strange Stories (lie, white lie, joke, pretence, misunderstanding, persuasion, appearance–reality, figure of speech, sarcasm, forgetting, double bluff, contrary emotion)	DLD and autistic groups gave fewer correct mental state answers than TD group, autistic group gave more inappropriate answers than DLD or TD
Rakhlin et al., 2011	Children with: DLD and IQ above 85 (*n* = 21), DLD and IQ below 85 (*n* = 4), IQ below 85 but no DLD (*n* = 5), TD (*n* = 22), aged 5;0–12;9, non-DLD groups older on average	8 story scenarios for FB: unexpected transfer	DLD groups performed worse than non-DLD groups; language development scores related to FB performance even after controlling for IQ and short-term memory
Spanoudis, 2016	Children with DLD (*n* = 20) aged 8;9–12;2, age- and gender-matched TD children (*n* = 20), gender- and language-matched controls (*n* = 18)	18 short stories incorporating faux pas recognition and the Strange Stories: 12 social indiscretions, 6 control stories	DLD group performed worse than age-matched controls; language and ToM skills were related: syntactic and pragmatic abilities predicted ToM performance

Table 2: Descriptive results of the TD and LT children's mean performance and score range across measures of language, metacognitive, and metalinguistic skills.

Test	**TD Mean (Score Range)**	**LT Mean (Score Range)**
Productive Vocabulary (TD *n* = 27, LT *n* = 8; T-Score)	54.44 (40–66)	48.25 (38–68)
Productive Grammar (TD *n* = 27, LT *n* = 8; T-Score)	60.44 (44–74)	55.38 (50–68)
Receptive Grammar (TD *n* = 27, LT *n* = 8; T-Score)	55.74 (46–69)	46.88 (37–53)
Reading Score (TD *n* = 26, LT *n* = 8; Reading Quotient T-Score)	47.12 (24.70–72.30)	44.84 (30.00–56.70)
ToM-Scale (TD *n* = 26, LT *n* = 8; maximum of 3)	2.12 (1–3)	2.00 (1–3)
Strange Stories (TD *n* = 26, LT *n* = 7; maximum of 16)	12.65 (9–16)	10.00 (4–13)
Metaphor Comprehension (TD *n* = 27, LT *n* = 8; maximum of 22)	10.07 (6–15)	7.00 (1–19)

Table 3: Background and performance of 8 children with a history of LT. a = Size of productive vocabulary measured with FRAKIS (Szagun et al., 2009); b = Standard T-values, measured with SETK-2 (Grimm, 2000); c = taking part in two units with a speech and language pathologist who explained and demonstrated language beneficial behaviour; d = did not participate at 5;0; LT = late talker; TD = typical development; DLD = developmental language disorder.

No.	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Sex	Male	Female	Female	Male	Female	Female	Male	Male
Family history of language delay	Yes	Yes	No	No	Yes	Yes	No	No
Size of productive vocabulary at 2;0[a]	23	69	123	95	41	152	28	3
Word comprehension at 2;0[b]	44	41	69	48	54	38	38	48
Sentence comprehension at 2;0[b]	26	26	35	35	54	26	35	54
Word production at 2;0[b]	26	32	33	32	30	36	30	26
Sentence production at 2;0[b]	30	34	Missing	34	35	36	39	30
Parental guidance[c]	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Status 2;6	LT	LT	LT	TD	TD	LT	TD	TD
Speech and language therapy	20 units	No	Breakup	---	---	10 units	---	---
Status 3;0-6;0	TD	DLD[d]	DLD	TD	TD	TD	TD	TD
Corollary diagnosis: dyslexia	No	No	No	Yes	No	No	No	No
Status 9;0	TD	TD	TD	TD	TD	TD	TD	TD
Productive vocabulary at 9;0	48	38	39	49	68	48	56	40
Productive grammar at 9;0	52	50	52	50	56	68	61	54
Receptive grammar at 9;0	49	46	37	49	53	42	46	53
Reading score at 9;0 (converted to T-score)	45.3	30.0	49.0	36.0	47.7	48.0	46.0	56.7
ToM-scale at 9;0	2	2	1	3	3	2	2	1
False belief	1	1	0	1	1	1	1	1
Hidden emotion	1	1	1	1	1	1	1	0
Sarcasm	0	0	0	1	1	0	0	0
Strange Stories at 9;0	8	11	4	13	12	11	11	Missing
Metaphor comprehension at 9;0	5	5	3	6	19	10	7	1

Table 4: Results of the Mann-Whitney-U-tests comparing language, metacognitive and metalinguistic skills between LTs and TD children including performance on each ToM-scale task. Bonferroni-type Benjamini and Hochberg (1995) corrections for multiple comparisons and controlling the false discovery rate are reported under *p**. For the highest level of accuracy, the original reading quotient for the reading score is used instead of our T-score conversion.

	TD		***LTs***						
	Mdn	***IQR***	***Mdn***	***IQR***	***U***	***Z***	***p***	***r***	***p*****
Productive Vocabulary (TD *n* = 27, LT *n* = 8)	55.0	10	48.0	15	55.0	-2.089	.037	.352	.065
Productive Grammar (TD *n* = 27, LT *n* = 8)	59.0	8	53.0	9	63.5	-1.762	.078	.296	.109
Receptive Grammar (TD *n* = 27, LT *n* = 8)	53.0	15	47.5	9	45.0	-2.514	.012	.418	.042
Reading Score (Original RQ) (TD *n* = 26, LT *n* = 8)	94.0	43.0	95.25	15.6	102.5	-.061	.951	.011	.951
ToM-Scale (TD *n* = 26, LT *n* = 8)	2.0	1	2.0	2	95.0	-.408	.683	.063	.797
Strange Stories (TD *n* = 26, LT *n* = 7)	13.0	3	11.0	4	38.5	-2.339	.019	.402	.044
Metaphor Comprehension (TD *n* = 27, LT *n* = 8)	10.0	3	5.5	6	44.5	-2.514	.012	.422	.042

Figure 1: Results of TD and LT children's performance and score range across the standardised measures of language and reading (T-scores on the y-axis, means, medians, inter quartile ranges, and outliers presented).