The Development of Prospective Processing of Simple and Complex Actions in Early Childhood

Inaugural-Dissertation
zur Erlangung des Doktorgrades der Philosophie
der Ludwig-Maximilians-Universität
München

vorgelegt von
Saskia Nicole Melzel (geb. Tobias)
aus Stuttgart
2022

Referent: Prof. Dr. Markus Paulus

Korreferentin: Prof. Dr. Beate Sodian

Tag der mündlichen Prüfung: 26.11.2021

Contents

Acknowledgements 1

Abstract 3

Zusammenfassung 7

1. General Introduction 11

1.1. Understanding Others' Actions: the Role of Gaze Measures 12

1.2. Action Anticipation: Simulation Theory vs. Lower-Level Mechanisms ... 16

1.3. Dual-Process Theories: Automatic and Controlled Processes 19

1.4. The Hierarchical Structure of Actions 23

2. The Current Thesis 29

2.1. Research Questions 29

2.2. Summary of the Studies and Author Contributions 31

3. General Discussion 37

3.1. Psychological Mechanisms Underlying Visual Anticipation 38

3.2. Automatic and Controlled Processes in Prospective Action Processing ... 41

3.3. Prospective Processing of Hierarchical Actions 45

3.4. Limitations and Directions for Future Research 46

3.5. Conclusion 50

4. References 51

5. Appendices 71

A) Study 1: The development of children's use of kinematic cues for action anticipation and action prediction 71

B) Study 2: The ontogeny of efficient second-order action planning: The developing interplay of controlled and habitual processes in goal-directed actions 125

C) Study 3a: The development of the prediction of complex actions in early childhood 127

D) Study 3b: The development of processing of nested structures in preschool children 129

Acknowledgements

Ohne die Unterstützung zahlreicher Personen wäre meine Dissertation in dieser Form nicht möglich gewesen. An dieser Stelle daher ein ganz herzliches *Dankeschön* an alle, die mich bei der Entstehung dieser Arbeit unterstützt haben.

Mein besonderer Dank gilt meinem Erstbetreuer Markus Paulus für die fachliche Expertise, das Vertrauen in meine Arbeit, die ansteckende Begeisterung für die Wissenschaft, die vielen kritischen Fragen und für das stete Verständnis, mit dem er meiner Arbeit begegnet ist. Ebenso gilt mein Dank Beate Sodian als Zweitgutachterin dieser Arbeit. Danke an Katrin Lindner für die Bereitschaft die Rolle der Drittprüferin zu übernehmen.

Danke an Angela Friederici und das MPI Leipzig für die Förderung meiner Arbeit.

Vielen lieben Dank an all die tollen KollegInnen, sowohl für den wertvollen fachlichen Austausch, als auch ganz allgemein für die schöne Zeit, auf die ich stets mit Freude zurückblicken werde. Insbesondere danke an Kerstin Ganglmayer, Marina Kammermeier, Samantha Lenz, Regina Sticker, Natalie Christner, Özgün Köksal, Tamara Haack, Nina Hinz, Samuel Essler, Carolina Pletti, Antonia Misch, Maria Mammen, Gökhan Gönül, Lena Söldner, Andrea Kramer, Katharina Heiß, Monika Wörle, Theresa Hagenauer und Anja Stadler.

Des Weiteren danke ich allen teilnehmenden Familien, sowie den studentischen Hilfskräften, insbesondere Sydney Rafael Häberer Haydar, Annika Moratzky, Laura Wenzlick, Nadine

Kühner und Annika Kaltenhauser für ihren tatkräftigen Einsatz bei der Datenerhebung. Danke an Petra Janßen für die große Hilfe bei jeglichen administrativen Angelegenheiten.

Zuletzt herzlichen Dank an Familie und Freunde für den stets liebevollen und ermutigenden Beistand. Insbesondere danke an meiner Schwester Yvonne Tobias-Miersch, die Teile der Dissertation gegengelesen hat und selbstverständlich meinen Eltern, die mich stets in meinem Weg bestärkt haben. Ein spezieller Dank gilt meinem Mann Roman Melzel für die unermüdliche Geduld und das immerwährende Verständnis.

Abstract

Everyday human life is characterized by social interactions, for which understanding and anticipating others' actions play a crucial role. Furthermore, planning one's movements in an anticipatory manner allows for acting smoothly and efficiently by avoiding time-consuming corrective movements. This thesis aimed to investigate how prospective processing of others' actions as well as of children's own actions develops over early childhood. One key focus was to examine the prospective processing of actions of varying degrees of complexity. Four empirical studies were conducted to expand our understanding of this topic by focusing on the following three aspects.

First, Study 1 examined two possible mechanisms underlying children's visual anticipations (as a key measure of prospective action understanding) during perceiving others' simple actions. It aimed at answering the question of whether simulation theories or lower-level perceptual mechanisms account for children's visual anticipations of others' simple grasping actions. Three-year-olds', 4-year-olds', 10-year olds', and adults' (*N* = 98) prospective action processing was assessed by their anticipatory gaze shifts within an eye-tracking study. Participants observed a hand reaching for one out of two objects while the motor kinematics of the reaching hand varied depending on whether it reached for a close or far object. Results revealed that none of the children age groups used the motor kinematics to correctly visually anticipate the target object. Instead, they showed a looking bias to the close object. This indicates that lower-level perceptual mechanisms (following the general movement of the reaching hand and being attracted by a salient object) and not motor simulation (matching the observed action onto their own motor repertoire) accounted for children's visual anticipations of simple other's simple grasping movements.

Second, the roles of automatic and controlled processes for children's prospective processing of others' actions as well as for their prospective action planning were examined. The results of Study 1 mentioned above indicated that children's visual anticipations were triggered rather automatically. The second experiment of Study 1 assessed participants' (*N* = 80) verbal predictions of the grasping actions described above. It turned out that 10-year-olds used kinematic cues to verbally predict the action target and that 4-year-olds learned to do so over the trials, whereas neither of the age groups correctly visually anticipated the action in the first experiment. These findings suggest that a more sophisticated understanding of other's actions might initially require explicit, controlled processing of the situation. This conflicts with the predominant view in developmental psychology that children show an implicit understanding of others' actions before showing an explicit understanding later in ontogeny. Furthermore, Study 2 examined the role of automatic and controlled processes for children's prospective action planning in a large sample (*N* = 246) of 2- to 14-year-old children and adults. Two conceptually different object manipulation tasks were assessed to investigate whether anticipatory movement planning develops as general capacity and to what extent participants rely on the habitual use of the object or plan their grasping movement in a controlled manner. The results suggest that anticipatory movement planning might develop as a general capacity. Furthermore, the findings are indicative of an increase of controlled processes for prospective action planning in early childhood, while the developmental pattern of the interplay of automatic and controlled processes was task-specific. In sum, Study 2 highlights the developmental dynamics of the interplay of habitual and controlled processes in children's own prospective action planning.

Third, Study 3a and 3b examined in more detail how the ability to prospectively process hierarchical actions develops in early childhood and whether it is related to children's processing of hierarchical structures in the language domain. In Study 3a, 3- to 6-year olds and adults (*N* = 111) were asked to verbally predict actions that required a means (sub-action) to achieve an actor's stated overarching goal. This task also served as one out of three complex action tasks in Study 3b, which investigated 3- to 6-year-old children's (*N* = 130) hierarchical structure processing across the action and language domain. Overall, the results suggest that children's ability to predict hierarchical actions increases over early childhood (Study 3a, 3b). Furthermore, Study 3b indicates that the processing of hierarchical structures might develop as a domain-general capacity and highlights the role of working memory for hierarchical structure processing.

Taken together, this thesis contributes to research on how young children prospectively understand others' actions and how they prospectively plan their own actions. First, it provides evidence that low-level perceptual mechanisms rather than motor simulation account for young children's visual anticipations of simple grasping actions (Study 1). Second, it points to substantial developments of controlled prospective processing over early childhood, which is supposed to be essential for processing more complex situations and actions (Study 1, 2, 3a, 3b). Lastly, children's development of prospectively processing hierarchical action and language structures seems related and might be explained by working memory (Study 3b).

Zusammenfassung

Unser tägliches Leben ist von sozialen Interaktionen geprägt. Die Handlungen anderer zu verstehen und antizipieren zu können ist für soziale Interaktionen von großer Bedeutung. Darüber hinaus ermöglicht eine antizipative Bewegungsplanung nahtlos ineinander übergehende und effiziente Handlungen ohne zeitaufwendige Korrekturbewegungen. Das Ziel dieser Dissertation ist zu untersuchen wie sich die prospektive Verarbeitung von Handlungen anderer Personen, als auch die eigene prospektive Handlungsplanung in der frühen Kindheit entwickelt. Ein zentraler Punkt der vorliegenden Arbeit ist, dass sie die prospektive Verarbeitung von Handlungen unterschiedlicher Komplexität betrachtet. Um unser Verständnis dieses Forschungsbereichs zu erweitern, wurden vier empirische Studien mit einem Fokus auf die folgenden drei Teilaspekte durchgeführt.

Erstens untersuchte Studie 1 zwei mögliche zugrunde liegende Mechanismen für visuelle Antizipationen (als ein zentrales Maß für prospektives Handlungsverständnis) von Kindern, während die Kinder eine einfache Handlung einer anderen Person beobachteten. Die Studie hatte zum Ziel herauszufinden, ob visuelle Antizipationen einfacher Greifbewegungen anderer durch Simulationstheorien oder durch Low-Level Wahrnehmungsmechanismen erklärt werden können. Die prospektive Handlungsverarbeitung 3-Jähriger, 4-Jähriger, 10-Jähriger und Erwachsener (*N* = 98) wurde anhand ihrer antizipatorischen Blickbewegungen im Rahmen einer Eye-Tracking Studie erfasst. Die Studienteilnehmer beobachteten, wie eine Hand nach einem von zwei Objekten griff. Die motorische Kinematik der greifenden Hand variierte in Abhängigkeit davon, ob ein näheres oder weiter entferntes Objekt gegriffen

wurde. Die Ergebnisse zeigten, dass keine der Kinder Altersgruppen die kinematischen Informationen nutzte, um das korrekte Zielobjekt zu antizipieren. Stattdessen antizipierten sie tendenziell das nähere Objekt, unabhängig von der motorischen Information. Dies deutet darauf hin, dass nicht die Simulation von Bewegungen (spiegeln einer beobachteten Handlung auf das eigene motorische Repertoire), sondern Low-Level Wahrnehmungsmechanismen (Antizipieren der Bewegungsrichtung der Hand und gleichzeitiges Angezogen werden von einem salienten Objekt) die visuellen Antizipationen der Kinder erklärten.

Zweitens wurde die Rolle automatischer und kontrollierter Prozesse für die prospektive Handlungsverarbeitung von Handlungen anderer Personen, als auch für die eigene prospektive Handlungsplanung in der frühen Kindheit untersucht. Die eben erwähnten Ergebnisse der ersten Studie wiesen darauf hin, dass die visuellen Antizipationen eher automatisch hervorgerufen wurden. In einem zweiten Experiment von Studie 1 sollten die Versuchspersonen (N = 80) die beschriebenen Greifbewegungen verbal vorhersagen. Es zeigte sich, dass 10-Jährige die kinematischen Informationen nutzten, um vorherzusagen welches der beiden Objekte gegriffen wird und dass 4-Jährige über die Trials hinweg lernten eine korrekte Vorhersage zu treffen, wohingegen keine der beiden Altersgruppen die Handlung im ersten Experiment korrekt antizipierte. Diese Befunde deuten darauf hin, dass ein differenzierteres Verständnis von Handlungen anderer Personen zunächst eine explizite, kontrollierte Verarbeitung der Situation erfordern kann. Dies steht im Konflikt mit der vorherrschenden Sichtweise in der Entwicklungspsychologie, dass Kinder ein implizites Verständnis von Handlungen anderer Personen haben, das sich erst später in der Ontogenese in ihrem expliziten Verständnis widerspiegelt. Darüber hinaus erforschte Studie 2 die Rolle automatischer und kontrollierter Prozesse für die eigene prospektive Handlungsplanung der Kinder in einer großen Stichprobe (N = 246) von 2- bis 14-jährigen Kindern und

Erwachsenen. Zwei konzeptionell unterschiedliche Objektmanipulationsaufgaben wurden erhoben. Einerseits, um herauszufinden, ob sich die antizipatorische Bewegungsplanung als eine allgemeine Fähigkeit entwickelt. Andererseits, um zu untersuchen, inwiefern das Greifen der Objekte von deren gewöhnlichen Gebrauch (automatisch) geleitet oder von kontrollierten Prozessen gesteuert wird. Die Ergebnisse deuten darauf hin, dass sich die antizipatorische Bewegungsplanung als allgemeine Fähigkeit entwickelt. Darüber hinaus lassen sie auf eine Zunahme von kontrollierten Prozessen für die prospektive Handlungsplanung in der frühen Kindheit schließen, während sich das Entwicklungsmuster des Zusammenspiels von automatischen und kontrollierten Prozessen als aufgabenspezifisch herausstellte. Zusammengefasst hebt Studie 2 die Dynamik der Entwicklung des Zusammenspiels von automatischen und kontrollierten Prozessen für die prospektive Handlungsplanung hervor.

Drittens wurde in den Studien 3a und 3b genauer untersucht, wie sich die Fähigkeit hierarchische Handlungen prospektiv zu verarbeiten in der frühen Kindheit entwickelt und ob sie mit der Entwicklung der Verarbeitung von hierarchische Sprachstrukturen zusammenhängt. In Studie 3a sollten 3- bis 6-jährige Kinder und Erwachsene (N = 111) Handlungen vorhersagen, die ein Mittel (eine Unterhandlung) erforderten, um ein angegebenes übergeordnetes Ziel eines Akteurs erreichen zu können. Diese Aufgabe diente auch als eine von drei hierarchischen Handlungsaufgaben in Studie 3b, in der die Verarbeitung hierarchischer Strukturen in 3- bis 6-jährigen Kindern (N = 130) in der Handlungs- und Sprachdomäne verglichen wurde. Zusammengefasst sprechen die Ergebnisse der Studien 3a und 3b dafür, dass die Fähigkeit hierarchische Handlungen vorherzusagen in der frühen Kindheit zunimmt. Ferner deuten die Ergebnisse darauf hin, dass sich die Verarbeitung hierarchischer Strukturen als domänenübergreifende Fähigkeit entwickelt und betonen die Rolle des Arbeitsgedächtnisses für die Verarbeitung hierarchischer Strukturen (Studie 3b).

Zusammengefasst trägt diese Dissertation zur Erforschung des prospektiven Handlungsverständnisses sowie der prospektiven Handlungsplanung in der frühen Kindheit bei. Erstens deuten die Ergebnisse darauf hin, dass Low-Level Wahrnehmungsmechanismen visuelle Antizipationen einfacher Greifhandlungen in der Kindheit erklären, während keine Evidenz für Simulationstheorien gefunden wurde (Studie 1). Zweitens weisen die Befunde auf eine erhebliche Entwicklung kontrollierter, prospektiver Verarbeitung von Handlungen in der frühen Kindheit hin, welche als essentiell angesehen wird, um komplexere Situationen und Handlungen zu verarbeiten (Studien 1, 2, 3a, 3b). Zuletzt scheint es einen Zusammenhang zwischen der Entwicklung der Verarbeitung von komplexen Handlungs- und Sprachstrukturen in der frühen Kindheit zu geben, welcher vermutlich durch das Arbeitsgedächtnis erklärt werden kann (Studie 3b).

1. General Introduction

Everyday human life is characterized by social interactions so that humans have commonly been referred to as “social animals” (e.g., Tomasello, 2014). Examples for social interactions range from sharing meals, over engaging in joint sports activities and playing games, up to driving a bike on a busy road. Being able to anticipate others’ actions as well as controlling one’s actions accordingly is crucial for social interactions, as merely reacting to others’ actions would not allow for smooth and fast action coordination (Sebanz & Knoblich, 2009). For example, anticipating others’ actions allows us to help others reaching their goal (e.g., if we see our partner reaching for the salt shaker out of his reach we pass it over), to protect others (e.g., by placing a hot cup of coffee out of a child’s reach), to protect ourselves (e.g., by anticipating that the car is not stopping at the pedestrian crossing), or to win games (e.g., by reading the opponent’s body language and jumping to the correct side of the goal). While these examples describe typical interactions of grown-ups, the question arises of how children become able to behave in such ways. More concrete, it raises the question about how children process others’ actions and about how children plan their own actions.

Therefore, the current thesis examines by means of four studies how young children prospectively plan their actions and how they prospectively understand others’ actions. One key focus of this thesis is how actions of varying complexity are processed in early childhood. This thesis aims at furthering our understanding of this topic by focusing on three different aspects. First, it examines two possible mechanisms underlying children’s visual anticipations (as a key measure of prospective action understanding) during perceiving others' simple actions. Second, it examines the role of automatic and controlled processes for children's prospective action planning as well as for their prospective understanding of other's actions. Third, it addresses the question of whether prospective processing of complex, hierarchical actions is related to children's processing of hierarchical structures in the language domain.

In the following parts, I will introduce the three questions that are at the center of the four empirical studies of this thesis. First, I will introduce a debate on the mechanisms underlying visual action anticipations. To embed this debate within the field of research, I will start by outlining the role of gaze measures to assess children's understanding of others' actions. Second, I will outline dual-process theories on prospective action processing and more precisely dual-process theories for social cognition and action control. Third, I will introduce the question of how children come to process more complex, hierarchical actions.

1.1. Understanding Others' Actions: the Role of Gaze Measures

How children process others' actions has commonly been investigated under the term action understanding. Noteworthy, action understanding has been used as an umbrella term, comprising various types of behaviors and competencies (for reviews see Thompson et al., 2019; Uithol & Paulus, 2014). It includes non-verbal and verbal aspects (often referred to as implicit and explicit understanding; Low & Perner, 2012) as well as actions of varying degrees of complexity, ranging from simple actions, such as reaching for or grasping an object, to more complex actions, such as baking a cake or planning a holiday trip. Distinct processes have been proposed to be involved in processing others' actions (Gredebäck & Daum, 2015). Whereas anticipating (non-verbal visual anticipation as well as verbal prediction) others' actions reflects prospective action processing, other facets of action understanding (e.g. imitation or looking time) assess an evaluation of the action typically after its' termination.

Examining children's action understanding using gaze measures has gained great interest. In contrast to imitation paradigms, gaze measures (such as looking time and visual anticipations) are not restricted by children's motor abilities. In other words, gaze measures allow for investigating children's processing of a great variety of other's actions, including actions which they are not yet able to perform themselves. Visual anticipations constitute a particularly interesting measure, as it allows for a direct examination of prospective action processing (in contrast to imitation or looking time measures). The methods to assess children's gaze underwent great technical progress in the last half-century and studies relying on gaze measures have contributed significantly to our today's understanding of how young children perceive others' actions. The next two sections will touch upon the history of looking time and visual anticipation measures and shortly discuss their contributions to the field of action understanding.

Looking Time

In a pivotal study, Fantz (1958) showed that the ability to discriminate patterns is present during the first six months of age. He assessed the relative length of infants' fixations to two simultaneously presented patterns. Infants showed a preference for one of the two patterns (by fixating it longer than the other pattern) if and only if the two patterns varied in type or degree of patterning. Consequently, the preferential looking indicated that infants were able to discriminate the presented patterns. This study laid the foundation for looking-time paradigms, such as visual habituation and violation of expectation paradigms, which played a crucial role in examining infants' action processing in the last two decades (e.g., Brandone & Wellman, 2009; Daum et al., 2009; Gergely et al., 1995; Woodward, 1998). For instance, in a well-known study by Woodward, 6-month-olds observed how a hand repeatedly reached for one of two objects (Woodward, 1998). After habituation, the positions of the two objects were swapped. Infants looked longer when the hand reached for the new object (old location), compared to when it reached for the object that has previously been grasped but was now in a new location, suggesting that they had encoded the relationship between the agent and the target. Since differences in looking time indicate which aspects of the observed action have been encoded, looking time studies are informative to whether children are sensitive to certain aspects of a previously observed action.

Noteworthy, some researchers have used the looking time paradigm to study infants' higher-level cognitive development. Resting on differences in looking time, it has for example been concluded that infants experience causality of events (Leslie & Keeble, 1987), that they evaluate the rationality of an agent's action (Gergely et al., 1995), and that they have arithmetic abilities (Wynn, 1992). However, such rich interpretations have been questioned on the ground that one needs to be very careful in linking looking time to underlying (hidden) constructs and that this linking needs to be done in the most conservative way possible (e.g., Aslin, 2007; Haith, 1998). One should bear in mind that the looking time paradigm has been developed to examine lower-level sensory and perceptual processes rather than high-level cognitive processing. As pointed out by Hunnius and Bekkering (2014), it can be very challenging to fully control for perceptual differences in the stimuli; longer looking times might thus often be better explained by low-level perceptual mechanisms rather than reflecting complex cognitive mechanisms in infants (Paulus, 2021; Ruffman et al., 2012; Ruffman, 2014).

Even though looking-time measures significantly contributed to getting deeper insights into infants' developing perception of others' actions, they reflect a retrospective evaluation of perceiving others' actions. Thus, they are inconclusive concerning children's prospective understanding of others' actions, which is the topic of the current thesis.

Visual Anticipations

In contrast to imitation or looking time measures, an assessment of anticipations allows for a direct examination of prospective action processing, that is, an examination of how actions are being processed over time while they are taking place (Gredebäck & Daum, 2015). That gaze patterns and particularly visual anticipations have gained so much interest in developmental psychology owes a lot to the technical progress of eye-tracking techniques. Early eye-tracking studies were characterized by quite invasive methods; Delabarre (1898) for example attached a wire ring with gypsum plaster to his cocainized eye to track his eye movements. He needed a recovery phase of a week before he could start the next testing. Yarbus (1967) refined the method and used a mirror that was attached to a kind of contact lens so that the reflection of a light source falling into the eye could be redirected and therefore externally recorded—a technique on which modern corneal reflection eye-trackers are still based on.

To extract relevant features from the environment saccades and fixations play a crucial role. Saccades are rapid, ballistic eye movements that allow bringing certain parts of the scene into the visual focus. This is essential as human vision is highly resolved only in a small part—the fovea centralis. Whereas visual perception is limited during saccades (saccadic suppression), the gaze remains on a single location during fixations, providing the basis for processing incoming visual information. As already pointed out by Yarbus, we "fixate on those elements of an object which carry or may carry essential and useful information" (Yarbus, 1967, p. 211). But how are essential elements selected? In the last 25 years, this question has been subject to many studies in the field of visual attention and led to heated debates about the extent to which visual selection is controlled by stimulus-driven bottom-up aspects or goal-driven top-down influences (see e.g., Theeuwes & Belopolsky, 2010 for a review). Leaving the discussion about the precise role of bottom-up and top-down aspects aside, it seems generally accepted that both play an important role in visual selection in adults.

To investigate prospective action processing, visual anticipations (i.e., predictive gaze shifts) have gained a lot of interest in developmental science—especially in research with preverbal children. Action anticipations are particularly interesting as they are informative concerning the observer's expectation about how the action unfolds. Noteworthy, anticipation studies suggest that children start to perceive others' actions as target-directed during the first year of life (e.g., Cannon & Woodward, 2012; Falck-Ytter et al., 2006; Hunnius & Bekkering, 2010; Kochukhova & Gredebäck, 2010). For example, in a study by Falck-Ytter et al. (2006) infants and adults observed how an actor repeatedly placed a toy into a bucket. In this study, 12-month-olds but not 6-month-olds showed predictive gaze shifts to the bucket in which the actor was about to place the toy. Other studies indicate that even 6-month-olds anticipate others' actions as target-directed (Hunnius & Bekkering, 2010; Kanakogi & Itakura, 2011; Kochukhova & Gredebäck, 2010). For example, Hunnius and Bekkering (2010) showed that 6-month-olds anticipate the mouth region of an actor when observing the actor lifting a cup.

Whereas these studies nicely demonstrate that infants anticipate the target location when observing simple actions, the underlying mechanisms of action anticipation are still under debate, as outlined in the next section.

1.2. Action Anticipation: Simulation Theory vs. Lower-Level Mechanisms

According to simulation theories, we understand and predict others' actions by simulating their mental states (e.g., Gallese et al., 2004; Goldman, 2006; Gordon, 1986; Rizzolatti et al., 2001). In general, simulation theories share the idea that we understand others by putting ourselves into their (mental) shoes (see Goldman, 2006; Gordon, 1986). In 1992 Di Pellegrino et al. discovered cells in the frontal area F5 of the macaque monkey which fired during the execution as well as during the observation of actions, so-called “mirror neurons”. This discovery subsequently led to considerations that the human motor system is involved in understanding others’ actions (for recent reviews see Heyes & Catmur, 2020; Thompson et al., 2019).

How the mirror neuron system might contribute to understanding others’ actions is still under debate. Some researchers suggest that it contributes to high-level action interpretation (e.g., Gallese et al., 2004; Rizzolatti & Sinigaglia, 2010). For instance, the direct-matching hypothesis claims that observing an action allows one to “understand directly the goal of the actions of others without needing inferential processing” (Rizzolatti & Sinigaglia, 2010, p. 268). Others suggest that the mirror neuron system rather contributes to the low-level processing of observed actions (Heyes & Catmur, 2020; Paulus, 2012). Besides these differences, it is usually assumed that observing an action can lead to an activation of the observer’s motor repertoire of the same action so that the observed action is “mirrored” in the observer’s own motor system and enables the observer to predict the upcoming action (Gallese et al., 1996; e.g., Gallese et al., 2004; Paulus, 2012; Rizzolatti et al., 2001). Indeed, there is compelling evidence indicating a close link between action perception and production in adults (Aglioti et al., 2008; Ambrosini et al., 2011; Fadiga et al., 1995; Sartori et al., 2011) and children (Ambrosini et al., 2013; Cannon et al., 2012; Daum et al., 2011; Kanakogi & Itakura, 2011; Kochukhova & Gredebäck, 2010; Paulus et al., 2012; Rosander & Hofsten, 2011; Sommerville et al., 2005; Southgate et al., 2009).

As our everyday actions are guided by pro-active gaze movements, the presence of anticipatory eye movements when observing someone else performing an action has been taken as evidence for simulation theories, and particularly for the direct-matching account (Falck-Ytter et al., 2006; Flanagan & Johansson, 2003; Kanakogi & Itakura, 2011). For example, Falck-Ytter et al. (2006) interpreted their finding that 12-month-olds and adults but not 6-month-olds anticipated to a bucket as support of the direct-matching account: as 6-month-olds (opposed to 12-month-olds and adults) are not able to master the observed action themselves, they cannot match the observed action on their motor repertoire and are thus not able to anticipate it. However, the role of the mirror neuron system for action anticipations has frequently been challenged (Hickok, 2014; Southgate, 2013).

Noteworthy, Falck-Ytter et al. (2006, p. 879) concluded that “when observing actions, 12-month-old infants focus on goals in the same way as adults do”. Relatedly, many other studies have stressed the role of goals[1] for action anticipations (Adam & Elsner, 2018; Cannon & Woodward, 2012; Falck-Ytter et al., 2006; Kanakogi & Itakura, 2011). Adapting the looking time paradigm from Woodward (1998), Cannon and Woodward (2012) showed 11-month-olds a hand grasping several times the same out of two objects. As in the original study, the positions of the two objects were then swapped and the hand started again to reach towards the objects, but stopped before indicating which of the objects she was going to grasp. They found that infants were more likely to show predictive gaze shifts to the old target (the one the hand has previously grasped) in the new location, than to the new target in the old location. This suggests that infants’ anticipations were based on information about the target rather than about the movement of the action. The authors concluded that “infants’ understanding of others’ goals […] shapes their online predictions about others’ next actions” (Cannon & Woodward, 2012, p. 297). However, comparable studies could not replicate the findings and instead highlighted the role of processing location-related over target-based information for action anticipations in early childhood (e.g., Daum et al., 2012; Ganglmayer et al., 2019). Noteworthy, one of the experiments was a direct but failed replication of the original study by Cannon and Woodward (2012) (Ganglmayer et al., 2019).

---

[1] It is important to note that the term goal has been used equivocally. As pointed out by Jacob (2009, p. 235): “it may refer to a physical target (e.g., a mug), to an act to be performed on a physical target (e.g., the agent’s bringing the mug to his mouth) and to an agent’s intention (e.g., the agent’s prior intention to drink).”

It has been suggested that action anticipations might often rather be explained by lower-level perceptual mechanisms, for example by learned visual associations (Heyes, 2014; Hunnius & Bekkering, 2010; Ruffman et al., 2012; Ruffman, 2014). The repeated co-occurrence of certain visual inputs might result in expectations of how the action unfolds. For example, Hunnius and Bekkering (2010) found 6-month-olds to anticipate the mouth region of an actor when observing the actor lifting a cup, providing evidence that infants have formed an association between the action of lifting a cup and the mouth region of a person. This lower-level approach is further supported by remarkable statistical learning abilities which are present even in very young children (Saffran & Kirkham, 2018). Whereas initial work on statistical learning focused on detecting regularities in continuous speech patterns (Saffran et al., 1996), recent research highlights the role of statistical learning for anticipating others' actions (Monroy, Gerson, & Hunnius, 2017; Monroy, Meyer, et al., 2017). This relates to claims that statistical learning abilities constitute a domain-general learning mechanism (Kirkham et al., 2002; Kirkham et al., 2007).

One contribution of this thesis is to progress this theoretical debate by directly examining in an empirical study whether simulation theories or lower-level perceptual mechanisms account for young children's visual anticipations when observing a simple grasping action (Study 1).

1.3. Dual-Process Theories: Automatic and Controlled Processes

As outlined in the previous sections, the use of non-verbal methods has revealed fascinating competencies in infants. This led to the question of how these early competencies can be explained in relation to children's later developing explicit reasoning about others' behavior (see Sodian et al., 2020 for an overview of various approaches). For example, it has been claimed that infants "analyze others' actions in terms of their intentional structure" (Cannon & Woodward, 2012, p. 292), whereas other studies indicate that children are not able to reliably verbally reason about the intention of others' actions until around 6 to 7 years of age (Bello et al., 2014). Relatedly, infants succeed in implicit non-verbal false-belief tasks during the second year of life (see Baillargeon et al., 2010; Sodian, 2016 for reviews), whereas they do not reliably pass explicit verbal false-belief task until around four years of age (Wellman et al., 2001). Many studies revealed comparable results, leading to the predominant view that children have an implicit understanding of others' actions, which is shown explicitly only later in development (Baillargeon et al., 2016).

However, as shown by recent empirical findings, children sometimes verbally predict others' actions before they correctly visually anticipate the same actions (Paulus et al., 2017; Schuwerk & Paulus, 2016), suggesting that an explicit understanding precedes their implicit understanding. For example, in a study by Schuwerk and Paulus (2016) an agent could either take a short or a long path to reach his goal. The agent announced that he intended to reach his goal as fast as possible. Whereas 5-year-olds were able to predict the agent's future action, they did not anticipate the correct path. This suggests that, depending on the action, children might first learn to reason about others' actions before being able to correctly anticipate them. Theories on skill acquisition suggest that a novel behavior (for example driving a car) is acquired by explicit training before it becomes an implicit, automatized skill (e.g., DeKeyser, 2015). In this sense, children might acquire the skill of correctly anticipating others' actions by rich experiences of talking and reasoning about others' actions (cf. Carpendale & Lewis, 2004).

Dual-process theories of social cognition

In contrast to accounts that propose a conceptual continuity between early implicit and later developing explicit mental state understanding (Baillargeon et al., 2016), the early implicit understanding of others' actions might be rather pre-conceptual, in line with dual-process theories of social cognition (e.g., Apperly & Butterfill, 2009; Lieberman, 2007; Strack &

Deutsch, 2004). According to dual-process theories, social cognition might be best explained by relying on two types of information processing systems: (1) an early developing, automatic, cognitively efficient but inflexible system on the one hand and (2) a later developing, cognitively demanding but flexible (controlled) system on the other hand. The automatic system is assumed to be based on rather simple associative processes, whereas the flexible system depends on language and executive functions, and allows for overcoming the limitations of the automatic system at the prize of cognitive effort. In adults, the two systems are supposed to exist in parallel (Apperly & Butterfill, 2009).

Whereas the flexible system is assessed by verbally reasoning about or predicting others' actions (e.g., Wellman et al., 2001), the implicit system is commonly assessed by gaze measures, such as visual anticipations (e.g., Barone et al., 2019). Previous research points to relations between children's early, non-verbal competencies of social cognition and their later explicit competencies (see Sodian et al., 2020 for a review). In contrast, a recent study suggests rather dissociated systems (Grosse Wiesmann et al., 2017). In the study by Grosse Wiesmann et al. (2017), implicit and explicit false-belief tasks were assessed in 3- and 4-year-olds and turned out to be not correlated. Moreover, performance in the explicit false-belief tasks—but not in the implicit false-belief tasks—was related to children's language and executive function abilities. The close relation between an explicit understanding of others' actions and developments in language and executive function is further supported by other studies (Baird & Astington, 2005; Devine & Hughes, 2014).

Taken together, given these ambiguities and diverging results, this thesis systematically investigates whether young children perform better in visually anticipating or verbally predicting others' actions when provided with kinematic cues of an actor, which indicate how a grasping action unfolds (Study 1).

Dual-process theories of action control

The difference between automatic and controlled processes has also been proposed in dual-process theories of action control (de Wit et al., 2012; de Wit & Dickinson, 2009; Hofmann et al., 2009; Wunsch & Weigelt, 2016). Whereas automatic processes are supposed to guide behavior from early on in life, controlled behavior is assumed to emerge later in ontogeny. Klossek et al. (2008) for example trained 18- to 48-month-olds to touch corresponding icons on a screen to view one of two types of video clips. Subsequently, participants were repeatedly exposed to one out of the two types of video clips, to cause a devaluation of this type. When children were allowed to touch the icons after devaluation, 2.5-year-olds but not 2-year-olds chose the icon that was associated with the still valuable video clips more often than the one associated with the devaluated clips. The authors concluded that younger children's response was probably elicited by the icon-outcome relation (i.e., icon → video) rather than by an expectation about an action-outcome relation (an action controlled in anticipation of the outcome: press certain icon → see still valuable video). This study indicates that controlling one's behavior with respect to anticipated action outcomes seems to develop after 2 years of age (for similar findings see Kenward et al., 2009).

On a theoretical level, it has been supposed that associations formed between perceptual inputs and behavioral outputs guide young children's behavior (e.g., de Wit & Dickinson, 2009). In other words, perceptual inputs are assumed to spontaneously and automatically evoke behavioral reactions. While such automatic processes are efficient concerning cognitive resources, they seem limited to overlearned (movement) patterns. Controlled processes, on the other hand, allow for overcoming such automatic, stimulus-driven reactions by planning one's behavior.

Relatedly, Karmiloff-Smith (1997a) has focused on the question of how children move from an implicit to an explicit action understanding. She argues that the child initially acquires procedural knowledge or action patterns, which are stored in rather inflexible, implicit, sequential representations. Such action patterns are supposed to allow for responding to stimuli in the external environment but the respective behavior is considered inflexible. After children have consistently mastered an action this “implicit information *in* the mind subsequently becomes explicit knowledge *to* the mind” by representational redescriptions and, thus, becomes more flexible and manipulable (Karmiloff-Smith, 1997a, p. 18). The redescription leads to overcoming the implicit, inflexible, independent representations and allows for drawing links between representations, for example by extracting common components to a new representation. Karmiloff-Smith considered the process of redescription to be uniquely human. According to her “nonhumans […] never become redescribers of the implicit knowledge embedded in their behavior, no matter how complex the behavior.” (Karmiloff-Smith, 1997b, p. 694).

Taken together, automatic and controlled processes are proposed to play a crucial role for prospectively controlling one’s actions as well as for understanding others’ actions. Given the theoretical considerations and empirical findings, it seems likely that young children initially rely on automatic processes, whereas controlled processes develop later in ontogeny. The current thesis capitalizes on these considerations by examining the role of automatic and controlled processes on children’s prospective action planning (Study 2) and understanding of others’ actions (Study 1, 3a, 3b).

1.4. The Hierarchical Structure of Actions

Noteworthy, a considerable amount of research has focused on children’s expectation about how simple actions (for example reaching or grasping actions) unfold over time (e.g., Adam et al., 2016; Cannon et al., 2012; Falck-Ytter et al., 2006; Hunnius & Bekkering, 2010; Kanakogi & Itakura, 2011; Rosander & Hofsten, 2011). While these actions are characterized by simple movements towards a target object, anticipations of such actions reflect expectations at which target location or target object an action is directed at. More generally, in simple actions, only the immediate upcoming action (opposed to a more distal action goal) is relevant for correct action anticipation or prediction.

However, everyday human behavior is more complex. We aim at achieving more abstract goals, which often require several actions to be fulfilled. So how is complex behavior organized? According to behaviorism, which dominated the first half of the 20th century, it was assumed that behavior can be explained sequentially by simple stimulus-response associations. Though, questioning such a linear chain structure of behavior, in which each element serves to arouse the next element by direct associations, Lashley (1951) (and later Miller et al., 1960) suggested that behavior is better explained as being hierarchically organized around action goals.

Flexibility has been put forward as one main advantage of hierarchically as opposed to sequentially organized behavior (see also Dawkins, 1976; Martins et al., 2019). Imagine you learn to play the piano. Given you are not very experienced, you repeatedly press the respective successive piano keys until you master the piece of music you want to play. Though, if you forget a note in the middle of the piece of music you just learned, you are probably not able to simply skip that note and continue playing. Instead, you likely have to start the sequence from the beginning (cf. Karmiloff-Smith, 1997a; Martins et al., 2019). Such sequential behavior has been explained by single action steps being only connected to their direct adjacent successor but not beyond so that if an action step fails, the sequence cannot be continued (Uddén et al., 2020).

It is assumed that hierarchically organized behavior allows us to overcome the limitations of sequential behavior, thus, paving the way for more flexible behavior. A frequently used example to demonstrate the hierarchical structure of behavior is the preparation of a cup of tea. For a cup of tea, one needs to put a tea bag in the cup and pour boiling water over it. In turn, putting a tea bag in the cup requires a cup in the first place (sub-action "get cup"), and pouring boiling water in the cup requires boiling the water (sub-action "switching on the kettle"). In other words, the preparation of a cup of tea comprises several actions, which can, in turn, comprise several sub-actions. Actions and sub-actions can be considered as means to achieve higher-level action goals, highlighting the means-end relations of hierarchically organized behavior (e.g., Botvinick, 2008). As behavior takes place in a dynamic environment, the actions necessary to achieve a certain goal vary depending on the situational constraints. That is, if the cup for my tea is in the dishwasher, I need to get it out of the dishwasher instead of the cupboard. Following this example, hierarchical behavior is organized around action goals of varying levels of abstraction, allowing for flexible behavior ranging from preparing a cup of tea, over planning a holiday trip, up to writing a doctoral thesis.

Different kinds of action hierarchies have been proposed, varying primarily in the assumed relations between the elements in the hierarchy (Uithol et al., 2012). According to the *action hierarchy,* actions are decomposed into sub-actions and sub-sub-actions, depicting a *part-whole relation* between the elements. For example, "preparing a cup of tea" consists of the sub-actions "put the tea bag in the cup" and "pour boiling water into the cup", whereas "put the tea bag in the cup" in turn consists of the sub-sub-actions "get the tea bag" and "get the cup". This kind of hierarchy *describes* an action according to a more abstract action goal. Another type of action hierarchy—the *control hierarchy*—focuses on the causal relations between the elements, thus, on controlling actions according to a goal. The idea of the control hierarchy is that higher-level elements (e.g., the desire for a cup of tea) can causally influence lower-level elements (e.g., getting a cup out of the cupboard) but not the other way around, depicting a top-down organization of action control. Whereas the *action* and *control hierarchies* are prevalent in action and motor control, Uithol et al. (2012) recently proposed an interesting third kind of action hierarchy: a hierarchy of *temporal extension*. In this type of hierarchy higher-level elements (goals) are represented longer and therefore influence an action over a longer time than lower-level elements (actions, motor acts). In the *temporal extension hierarchy,* elements of all levels can influence each other, broadening the strict top-down causal influence of the *control hierarchy*.

Besides these theoretical considerations, first empirical evidence indicates that the processing of more complex actions significantly improves over preschool years (Bello et al., 2014; Flynn & Whiten, 2008; Freier et al., 2017). For example, Freier et al. (2017) instructed 3- and 5-year-olds to color in six shapes (low-level action) according to an overarching goal (using each of three crayons equally often). Both, 3- and 5-year-olds showed good performance at the lower level of coloring-in the shapes, but only 5-year-olds consistently aligned their coloring-in activity to the higher-level action goal. This might indicate that 3-year-olds rather relied on their procedural knowledge of the coloring-in activity, whereas 5-year-olds more flexibly aligned their action of coloring-in to the higher-level action goal of using each crayon equally often. Whereas the study of Freier et al. (2017) examined children's action control according to a higher-level goal, Bello et al. (2014) investigated children's understanding of others' higher-level action goals (intentions). They showed that 3-year-olds were able to identify simple actions (such as grasping and touching) when they observed pictures of hand-object interactions. However, only 6- to 7-year-olds were able to reason about why an object was being grasped (e.g., to be placed somewhere else vs. to be used), that is, showing an understanding of the hierarchically higher goal of the simple motor act.

Strikingly, of the few studies that investigated children's understanding of complex actions, most relied on paradigms in which children were asked to evaluate an action after having observed it, that is, retrospectively (Bello et al., 2014; Flynn & Whiten, 2008; Whiten et al., 2006). However, little is known about children's prospective understanding of others' more complex actions. This is unfortunate, as human actions are often rather complex so that the hitherto existing focus on prospective processing of simple actions (Adam & Elsner, 2018; e.g., Cannon & Woodward, 2012; Daum et al., 2011; Daum et al., 2012; Elsner & Adam, 2020; Falck-Ytter et al., 2006; Ganglmayer et al., 2019) has limited developmental theorizing. Therefore, one contribution of this thesis is to empirically examine the development of young children's prospective action processing of more complex, hierarchical actions (Study 3a, 3b).

Hierarchical structures have been proposed to also play an essential role in other domains—most prominently in the language domain (Chomsky, 1956; Friederici et al., 2011; Greenfield, 1991). It is widely accepted that the syntactical structure underlying language, that is, the combination of several words (lower-level units) into phrases and sentences (higher-order elements), is hierarchical. Whereas researchers seem to agree that both action and language show a hierarchical organization of elements—as opposed to simple linear chains—the question of whether hierarchical structures are comparable across the action and language domain is subject to a recent controversial debate. Whereas some researchers highlight the similarities of hierarchical structures across the language and action domain (Pastra & Aloimonos, 2012; Pulvermüller, 2014), and suggest that hierarchical structure processing constitutes a domain-general capacity of human cognition (Fadiga et al., 2009; Grossman, 1980; Lashley, 1951; Marcus, 2006), others question the comparability of hierarchical structures across the domains (e.g., Zaccarella et al., 2021).

Notwithstanding the theoretical interest in hierarchical structures, little is known about how the ability to process hierarchical structures develops in early childhood. This thesis makes a novel contribution to this debate by systematically examining whether the development of hierarchical structure processing is related across the action and language domain and to which extent this relation depends on more basic cognitive capacities, such as working memory or inhibitory control (Study 3b).

2. The Current Thesis

Taken together, the current thesis aims to explore how young children prospectively process simple and complex actions. It focuses on the development of prospective action processing for children's movement planning as well as for their understanding of others' actions.

2.1. Research Questions

Within the presented field of research, this thesis investigates the following main research questions:

1. The first question relates to two possible mechanisms underlying children's prospective action processing—here: visual anticipations—of others' simple actions. Do simulation theories account for young children's visual anticipations of others' simple grasping actions or can their anticipations rather be explained by lower-level perceptual mechanisms? Whereas previous anticipation studies have been taken as evidence for simulation theories (e.g., Falck-Ytter et al., 2006), lower-level mechanisms cannot be ruled out as an alternative explanation for the results.
2. The following questions share a focus on the role of automatic and controlled processes in children's prospective action processing:
    a. The first sub-question relates to children's prospective processing of others' actions and in particular to their early competencies in implicit tasks compared to their later competencies in explicit tasks. Here, I investigated and systematically compared two aspects of prospective

    action processing: visual anticipation and verbal prediction of actions. Do visual anticipations reflect an understanding of others' actions that is shown explicitly only later in ontogeny (being the predominant view in developmental science)? Or do visual anticipations rather rely on automatic processes whereas more complex situations might require an explicit processing of others' actions, in line with dual-system theories of social information processing (e.g., Apperly & Butterfill, 2009)?

    b. The second sub-question relates to children's own prospective action planning: How does prospective movement planning develop in early childhood? More precisely, does anticipatory movement planning develops as a general capacity or is it task-specific? According to dual-process theories of action control, behavior relies on rather automatic processes on the one hand and controlled processes, which allow overcoming automatic reactions to perceptual inputs, on the other hand (de Wit & Dickinson, 2009; Hofmann et al., 2009). Whereas first empirical evidence indicates an interaction of habitual and controlled processes in 3- to 5-year-olds' motor planning for a specific type of tasks (Jovanovic & Schwarzer, 2017), I investigate how the interplay of habitual and controlled processes unfolds over development and whether the interaction should be considered task-specific or general.

3. The third research question focuses on whether the prospective processing of hierarchical action structures and hierarchical language structures show similar developmental pathways. More precisely: does the processing of hierarchical structures develop as domain-general capacity across language and action? Here, I

systematically examining whether the development of hierarchical structure processing in early childhood is related across the action and language domain.

2.2. Summary of the Studies and Author Contributions

To address these questions, four studies were conducted, assessing several age groups ranging from toddlerhood to adulthood. The next paragraphs will give a summary of the studies and relate them to the research questions. Furthermore, my contributions to each of the four studies are depicted in Table 1.

Study 1 served to examine children's prospective processing of others' simple actions. More precisely, it aimed at addressing the first research question as well as the first sub-question of the second research question. To this end, two experiments were conducted with 3-, 4-, 10-year-olds, and adults. The setup was adapted from the study of Falck-Ytter et al. (2006), given its prominent support for motor simulation theories. In both experiments participants observed a hand reaching for one out of two objects, both being located in the movement direction of the reaching hand (one closer one further away). The motor kinematics of the reaching hand varied depending on whether it reached for the close or far object. The two experiments assessed participants' expectations about which of the two cubes the hand was going to grasp: The first experiment was an eye-tracking study and assessed participants' ($N = 98$) visual anticipations and the second experiment assessed participants' ($N = 80$) verbal predictions. Regarding the first research question, the condition in which the hand reached for the far object allowed for differentiating between simulation and low-level perceptual accounts. According to simulation accounts, one would expect all age groups to be able to use the kinematic cues to correctly anticipate the target object, as the observed actions can be assumed to be part of participants' motor repertoire. If, on the other hand, lower-level perceptual mechanisms account for anticipations, participants were expected to show a preference for the close target, irrespective of the kinematic cues. That is, they were expected to anticipate the general movement direction of the reaching hand until coming across the next salient object. The results show that only adults, but neither of the children age groups based their anticipations on the kinematic cues. Instead, all children age groups showed a looking bias towards the close object, irrespective of the kinematics. Concerning the second research question, I investigated whether children anticipate the reaching action before being able to verbally predict it (in line with the predominant "implicit understanding precedes explicit understanding" view). If, however, the use of kinematic cues initially requires explicit processing and if kinematic cues can be processed automatically only later in ontogeny, we expected children to be able to verbally predict the action before being able to visually anticipate it. Experiment 2 shows that 10-year-olds used the kinematic cues to verbally predict the action and 4-year-olds learned to do so over the trials, whereas neither of the age groups was able to use the kinematic cues to visually anticipate the actions in the first experiment, thus supporting the second account. Together, the findings indicate that visual anticipations rather rely on automatic processes, whereas more complex situations require an explicit processing of others' actions. This is in line with dual-system theories of social information processing (e.g., Apperly & Butterfill, 2009).

Study 2 served to examine how prospective action planning develops over early childhood (Research Question 2b). Acting efficiently plays an important role in our everyday life. Planning movements in an anticipatory manner allows for smooth movement transitions and prevents us from time-consuming corrective movements. Until today, there is no conclusive evidence whether young children's efficient motor planning is task-specific or whether children develop a general capacity for efficient motor planning. Furthermore, current theories propose two types of processes to be involved in action control: largely automatic processes on the one hand and controlled, goal-directed processes on the other hand (de Wit et al., 2012; de Wit & Dickinson, 2009; Hofmann et al., 2009). How the interplay of such automatic processes and controlled processes unfolds over development has scarcely been investigated. Study 2 aims at closing this gap, by investigating efficient motor planning in 2- to 14-year-old children and adults (*N* = 246) in two conceptually different object manipulation tasks, using everyday objects (a spoon and a cup). To examine to what extent participants planned their grasping movement in a controlled manner or relied instead on the habitual use of the object, the orientation of the object was manipulated in such a way that either a habitually congruent or incongruent grasp was required for acting efficiently. The results indicate that children develop a general capacity for anticipatory movement planning. Furthermore, the results point to an increase of controlled processes for movement planning over early childhood. The developmental pattern of the interplay of automatic and controlled processes turned out to be task-specific and was not suggestive for a simple linear developmental trend. Taken together, the study highlights the developmental dynamics of the interplay of controlled and habitual processes in goal-directed action control and more generally expands our knowledge on the ontogeny of prospective motor planning.

Study 3a served as a pre-study for Study 3b and investigated the development of young children's prediction of others' complex actions (Research Question 3). As pointed out in the introduction, everyday behavior is more complex than simple reaching movements and has been proposed to be organized hierarchically (Lashley, 1951; Miller et al., 1960). Whereas a considerable amount of research has focused on the development of anticipating simple actions (e.g., Cannon & Woodward, 2012; Falck-Ytter et al., 2006), little is known about the development of prospective processing of more complex, hierarchical actions. Study 3a aimed at closing this gap by investigating the development of complex action prediction in 3- to 6-year-olds (*N* = 86) and an adult control group (*N* = 25). We relied on a paradigm that has previously been used to examine action prediction in young children (Hofsten et al., 2007; Paulus et al., 2017). Participants were asked to predict actions of two different degrees of complexity: in simple actions, an actor's goal could be achieved directly (by a single action), whereas, in complex actions, an additional action (means) was required to achieve an actor's goal. The results show that children's ability to predict simple and complex actions increases over early childhood. Even though children performed overall above chance when predicting complex actions, they performed significantly worse when predicting complex compared to simple actions. The adult control group showed ceiling performance in both conditions. In sum, the results show that children's prospective processing of complex, hierarchical actions shows substantial development over preschool years. Study 3a expands previous research that investigated complex action processing retrospectively (e.g., Bello et al., 2014; Freier et al., 2015).

Study 3b examined the processing of hierarchical structures across the action and language domain. More precisely, it aimed at investigating how the ability to process hierarchical structures develops in early childhood (Research Question 3). To examine whether it develops as domain-general capacity, as has been proposed by influential theories (Fadiga et al., 2009; Grossman, 1980; Jeon, 2014; Lashley, 1951; Marcus, 2006), hierarchical structure processing was assessed in three- to six-year-old children (*N* = 130) across two domains, action and language. To assess hierarchical structure processing in the action domain, three different types of tasks were used: a prospective action prediction task (see the complex condition of Study 3a), a task requiring controlled action (adapted from a study by Freier et al., 2017), and a task in which children had to copy a hierarchical structure (adopted from the prominent study by Greenfield & Schneider, 1977). The language tasks comprised two picture matching tasks, which assessed children's understanding of single and double embedded relative clauses (following a study by Fengler et al., 2016). Furthermore, Study 3b examined whether more general cognitive abilities (inhibitory control, working memory) account for children's capacity to process hierarchical structures and whether hierarchical structure processing contributes to the early development of Theory of Mind (ToM). Study 3b has three key findings: First, the results show a correlation of hierarchical structure processing across the action and language domain in early childhood. Second, the correlation was not significant anymore when controlling for working memory, but remained when controlling for inhibitory control. Third, hierarchical structure processing explained age effects in ToM and predicted ToM even after controlling for inhibitory control and working memory. Taken together, the findings hint at the development of a domain-general capacity of hierarchical structure processing across the action and language domain in early childhood and highlight the role of working memory for hierarchical structure processing. Consequently, Study 3b expands previous research which investigated the development of complex structure processing in early childhood within the language domain (e.g., Fengler et al., 2015) and action domain (Bello et al., 2014; Whiten et al., 2006) separately, by a direct comparison between the domains. Furthermore, the results support theories suggesting that hierarchical processing plays an important role in the early development of ToM (Frye et al., 1995).

Taken together, the studies of the current thesis provide evidence that prospective action processing shows substantial development over early childhood. It indicates that visually anticipating others' simple actions is subject to rather automatic and inflexible processes (at least in early and late childhood), whereas no evidence was found for simulation accounts (Study 1). Furthermore, more complex situations required (at least initially) explicit, controlled processes (Study 1). Concerning the development of children's own action planning abilities, the results indicate that that controlled movement planning increases over early childhood and might develop as general capacity while showing complex developmental patterns of the interplay of automatic and controlled processes (Study 2). Lastly, Study 3a demonstrated that prospectively predicting others' complex actions significantly improves over early childhood and that the development of processing complex, hierarchical structures is related across the action and language domain (Study 3b).

Table 1
Overview of the author's contributions to each of the four studies.

	Study 1	**Study 2**	**Study 3a**	**Study 3b**
Design			✓	✓
Data collection			✓	✓
Data analysis	✓	✓	✓	✓
Manuscript preparation	✓	✓	✓	✓

3. General Discussion

Prospective action processing plays a key role for smooth and fast action coordination when interacting with others (Sebanz & Knoblich, 2009), as well as for planning one's own actions (Rosenbaum, 2009). Merely reacting to others' actions or not planning one's actions can even be dangerous. For example, if you see a child reaching towards the hot stove (the child is not prospectively planning its action), you want to intervene before the child gets hurt (therefore, you need to be able to anticipate that the child is about to touch the stove). Furthermore, everyday behavior is more complex than simple reaching movements. It aims at reaching more abstract goals and has been proposed to be hierarchically organized (e.g., Lashley, 1951; Miller et al., 1960). So how do children in early childhood process others' actions and do they prospectively plan their own actions?

To enhance our understanding of this topic, the current thesis investigated the development of prospective action processing in early childhood by the means of four empirical studies and focused on three aspects: First, it investigated two possible mechanisms underlying children's visual anticipations during perceiving others' simple actions (Study 1). Second, it examined the role of automatic and controlled processes for children's prospective processing of other's actions and their prospective action planning (Study 1 and 2). Third, it investigated whether the processing of hierarchical action structures relates to hierarchical language structure processing (Study 3a and 3b).

In the following, the contributions of this thesis to developmental research on prospective action processing will be outlined in more detail, followed by a discussion of the limitations and suggestions for future research.

3.1. Psychological Mechanisms Underlying Visual Anticipation

The results of Study 1 indicate that visual anticipations of simple grasping actions in early and late childhood can rather be explained by lower-level perceptual mechanisms, whereas there was no evidence for simulation theories (Research Question 1). The findings suggest that children anticipated the general movement direction of the hand, while their gaze was attracted by the most salient region (here: closer object, more salient due to its greater proximity to the hand), which was then brought into the highly resolved focus of vision. Note, that the relevance of object saliency for visual anticipations in early childhood has also been highlighted in other studies (Adam et al., 2016; Henrichs et al., 2012). Consequently, the results of Study 1 question interpretations of previous anticipation studies, which were taken as support for simulation theories, in particular studies involving a single highly salient target object. For example, Falck-Ytter et al. (2006) interpreted their finding that 12-months-olds but not 6-months anticipated a bucket in which an actor was about to place a toy as support for simulation accounts, as 12-month-olds but not 6-month-olds master this action themselves. Though, considering the findings of Study 1 it seems more likely that 12-month-olds' gaze shift to the bucket was simply triggered by its highly salient nature (a red bucket with a 3D happy face connected to it).

Relatedly, it seems widely accepted that infants anticipations reflect an understanding of others' higher-level action goals (Cannon et al., 2012; e.g., Cannon & Woodward, 2012; Falck-Ytter et al., 2006; Woodward et al., 2009) and it has been suggested that simulating others' might support this ability (Woodward & Gerson, 2014). A prominent study that has significantly contributed to this view is the one by Cannon and Woodward (2012). They noticed that the results of Falck-Ytter et al. (2006) might also be explained by anticipations of the movement, being inconclusive to whether infants indeed anticipated the goal of the action. To overcome this limitation they used two objects at two locations. They showed that infants were more likely to base their anticipations on information about the target than the movement of the action and concluded that "infants' understanding of others' goals […] shapes their on-line predictions about others' next actions" (Cannon & Woodward, 2012, p. 297). However, comparable studies could not replicate the findings and highlighted the role of processing location-related over target-based information for action anticipations in early childhood (e.g., Daum et al., 2012; Ganglmayer et al., 2019). Noteworthy, one of the experiments was a direct but failed replication of the original study by Cannon and Woodward (2012) (Ganglmayer et al., 2019). The results of Study 1 support the findings that young children rely on the movement information when anticipating others' simple grasping or reaching actions. No evidence could be found that they base their anticipations on the higher-level goal of the action (grasping the close vs. far cube) by matching the action onto their motor repertoire, challenging direct matching accounts for visual anticipations.

These considerations seem also interesting in the light of prominent developmental theories. Jean Piaget's pioneering work has suggested that thinking in the preoperational phase (2-7 years of age) is characterized by egocentrism (that is, the tendency to perceive the word from one's own point of view) (Siegler et al., 2016). This suggests that young children might process others' behavior rather from their *own point of view* (that is, with respect to the behavior they observe) than with respect to *others' goals, beliefs, or desires* (but see Baillargeon et al., 2010; Baillargeon et al., 2016).

Study 1 furthermore supports claims that learned visual associations play a key role for prospective gaze measures (Heyes, 2014; Hunnius & Bekkering, 2010; Ruffman et al., 2012; Ruffman, 2014). Assessing children's looking time to either of the two objects before the hand reappeared behind the occluder (i.e., prospectively) indicated a learning effect in 10-year-olds over the trials. Note that the learning effects were not examined for first fixations but only for the prospective looking time. Though, the remaining analyses point to similar results for the two prospective action processing measures. The learning effect indicates that visual associations formed by the repeated co-occurrence of the motor kinematics of the specific grasping (that is, reaching for the close vs. far object) and the grasping of the respective object might eventually allow for correct action anticipation. This is in line with children's remarkable statistical learning abilities, which are present already early in life (Saffran & Kirkham, 2018). Indeed, the role of statistical learning for anticipating others' actions has been highlighted recently (e.g., Monroy, Gerson, & Hunnius, 2017). Study 1 supports views that perceptual, associative learning is important for visual anticipations in early childhood.

Whereas Study 1 could not find any evidence that simulation theories account for young children's visual anticipations of others' simple reaching actions, the question remains how the close link between action perception and production in childhood can be explained (Cannon et al., 2012; Daum et al., 2011; Kanakogi & Itakura, 2011; Kochukhova & Gredebäck, 2010; Paulus et al., 2012; Rosander & Hofsten, 2011; Sommerville et al., 2005). One possibility is that motor activation *follows* action identification when observing others' actions rather than being a prerequisite for action understanding. In other words, motor activation might rather be a result than a contributor to action identification, consequently questioning its role for understanding others' actions. First evidence for this alternative view comes from a study with adults (Pomiechowska & Csibra, 2017). Future research might elaborate on this alternative view and investigate it in a children sample. Whereas previous studies have shown that children's action experience alters their perception of others' actions (Daum et al., 2011; Kochukhova & Gredebäck, 2010; Sommerville et al., 2005; Sommerville et al., 2008), the findings are inconclusive with respect to the exact role of the motor activations.

In sum, this thesis makes a novel contribution to previous research by systematically investigating whether simulation theories or lower-level perceptual mechanisms account for young children's visual anticipations when observing a simple grasping action (Study 1). The results indicate that children's anticipations were based on lower-level perceptual mechanisms, whereas no evidence was found for simulation accounts.

3.2. Automatic and Controlled Processes in Prospective Action Processing

The thesis stresses the role of automatic processes for prospective action processing from early on in life and points to substantial developments of controlled processes over early childhood. As will be discussed in more detail in the following paragraphs, the development of controlled action processing is reflected in children's prospective processing of others' actions (Study 1, Study 3a, Study 3b) as well as in their action planning and execution (Study 2, Study 3b).

As outlined in section 3.1, children did not base their visual anticipations flexibly on the motor kinematics (Study 1). Instead, their anticipations were directed at the first salient object in the general movement direction, irrespective of the kinematic cues. This supports previous considerations and findings, highlighting the inflexibility of visual action anticipations in early childhood (e.g., Daum et al., 2012; Ganglmayer et al., 2019; Ruffman et al., 2012).

Interestingly, Study 1 revealed clear discrepancies between children's visual anticipations and verbal predictions (Research Question 2a). Note that whereas 10-year-olds used the kinematic cues to correctly verbally predict the action, and 4-year-olds learned to do so over the trials, none of the children age groups used the cues to correctly visually anticipate the action. The results of Study 1 are thus inconsistent with the predominant view that children understand others' actions implicitly before showing an explicit understanding later in ontogeny (e.g., Baillargeon et al., 2016). Instead, Study 1 adds to recent empirical findings demonstrating that action prediction sometimes precedes action anticipation in early childhood (Paulus et al., 2017; Schuwerk & Paulus, 2016). This indicates that in more complex situations children might have to learn to reason about others' behavior first (relying on explicit, controlled processes) before this knowledge is transferred (by training) into an implicit, automatized skill of correctly anticipating others' behavior (cf. Carpendale & Lewis, 2004). In this sense, visual anticipation of others' more complex behavior might be considered as a to be acquired skill (see DeKeyser, 2015 for a theory on skill acquisition).

Overall, the thesis points to substantial developments of controlled processes for prospective action processing in early childhood. Even though Study 1 was not specifically designed to investigate the precise developmental pattern of controlled prospective processing of others' actions over early childhood, the results indicate that children's reasoning abilities about others' behavior (relying on controlled processes) are subject to significant improvements over early and middle childhood. Whereas 10-year-olds were able to use the kinematic cues to correctly verbally predict the action, and 4-year-olds learned to do so over the trials, 3-year-olds performed at chance. The growth of controlled processes for a prospective understanding of others' actions is further supported by the results of Study 3a. Predicting other's complex actions requires planning the action with respect to the specified higher-level action goal, that is, in a controlled manner. As shown in Study 3a, children's ability to correctly predict an actor's complex action significantly improved over early childhood.

Further support for an increase of controlled processes over early childhood comes from Study 2 and 3b (coloring-in task), which assessed children's own prospective action planning and execution abilities. Study 2 revealed an increase of controlled action planning over early childhood (Research Question 2b): Children increasingly planned their movement with respect to a given action outcome instead of relying on the grasp associated with the habitual use of the object. Moreover, the developmental pattern of the interaction of controlled and automatic processes in Study 2 suggests that automatic behavior is not simply replaced by controlled behavior. Instead, the pattern is suggestive for a more complex interplay of automatic and controlled processes for prospective movement planning over development and this interplay seems to depend on the respective action context. In a similar vein, the results of the coloring-in task of Study 3b suggest that whereas children around the age of three perform well in routinized, habitual behavior (i.e., engaging in a coloring-in activity), they show significant improvements in aligning their actions with respect to a given higher-level action goal over early childhood (here: using each crayon equally often)[2].

Noteworthy, previous research on the development of prospective movement planning points to substantial developments during early and middle childhood (for reviews see Pereira et al., 2019; Wunsch et al., 2013). Though, as previous studies usually assessed children's prospective motor planning in a single task (e.g., Adalbjornsson et al., 2008; Ansuini et al., 2018; Scharoun, Robinson, et al., 2018; Thibaut & Toussaint, 2010; Weigelt & Schack, 2010) or compared children's mean performance across two tasks (Knudsen et al., 2012; Scharoun Benson et al., 2018), there is no conclusive evidence whether prospective motor planning develops as general capacity. Given that children's performance varied across studies and were dependent on the tasks applied, children's tendency to show prospective motor planning has been claimed to be task-specific (see Rosenbaum et al., 2012; Wunsch et al., 2013).
Study 2 extends this line of research by systematically investigating whether young children develop a general capacity for efficient motor planning. The results show a relation between

---

[2] Note that additional correlational analyses (not reported in Study 3b) between children's age and their task performance in the coloring-in task revealed similar findings as the original study by Freier et al. (2017).

two conceptually different movement planning tasks. This hints at the possibility that prospective motor planning might involve the development of a general capacity (Research Question 2b).

That prospective movement planning might develop as general capacity (Study 2) is also interesting concerning the development of prospective processing of hierarchically structured actions. Participants' grasping in Study 2 is regarded as prospectively planned if the second action (insertion of the object) is taken into account during the planning of the first action (grasping of the object). This behavior can be considered as hierarchically organized, as a simple sequence of actions, in which each element serves to arouse the next element by direct associations, seems insufficient to explain the influence of the second action on the first action. It might be speculated that a general capacity to process more complex structures underlies children's developing prospective motor planning (see also section 3.3 and 3.4).

In sum, this thesis highlights the interplay of automatic and controlled processes for prospective action processing. It points to substantial developments of controlled processes over early childhood for prospective processing of (1) others' actions as well as for (2) children's action planning. Overall, the findings are in line with dual-system theories of social cognition (e.g., Apperly & Butterfill, 2009) and human action control (e.g., de Wit & Dickinson, 2009). The thesis provides new empirical evidence that more complex action contexts might initially require explicit, controlled processing of the situation (Study 1). This seems particularly relevant considering ongoing debates about the relation between children's implicit and explicit action understanding (e.g., Baillargeon et al., 2016; Ruffman, 2014). Furthermore, the thesis provides first evidence that controlled motor planning might develop as a general capacity (Study 2). This expands previous research, which has been inconclusive concerning this point.

3.3. Prospective Processing of Hierarchical Actions

This thesis indicates that children's prospective action processing is increasingly aligned towards more distal action outcomes (Study 2, 3a, and 3b). As shown by Study 2, children start to take a future action into account, when planning their first action and the results indicate that this might develop as general capacity rather than being task-specific. Furthermore, when predicting others' actions as well as during their action execution they increasingly consider the means-end structure of the action (Study 3a and 3b), which has been proposed to be crucial for hierarchically organized behavior (e.g., Botvinick, 2008). These results are particularly interesting, as they make a new contribution to previous research, which primarily investigated children's hierarchical action understanding after they have observed an action (e.g., Bello et al., 2014; Flynn & Whiten, 2008; Whiten et al., 2006).

Noteworthy, the results of Study 3b are suggestive for the development of a domain-general capacity of hierarchical structure processing across the action and language domain (Research Question 3). The development of a general capacity of prospective hierarchical structure processing is further supported by Study 2, which revealed that children's own prospective movement planning was related across two conceptually different tasks. Whereas previous research pointed to significant improvements of hierarchical structure processing within the language (Corrêa, 1995; Fengler et al., 2016; Villiers et al., 1979) and action domain (Flynn & Whiten, 2008; Freier et al., 2015, 2017; Greenfield & Schneider, 1977; Pereira et al., 2019; Rosenbaum et al., 2012), the current thesis expands this research by presenting first empirical evidence that hierarchical structure processing in young children might be related across the domains. This seems of particular relevance, considering recent theoretical discussions about the comparability of hierarchical structures across the action and language domain (e.g., Pastra & Aloimonos, 2012; Pulvermüller, 2014; Zaccarella et al., 2021).

Interestingly, working memory accounted for the relation between the two domains. This adds to previous empirical research, which highlighted the role of working memory for the processing of hierarchical language structures (e.g., Fengler et al., 2015; Meyer et al., 2013), and suggests a more general role of working memory for the development of processing hierarchical structures.

Why is working memory so important for processing hierarchical structures? And more specifically, what role does it play for prospective processing of hierarchical actions? One might assume that working memory allows for maintaining and coordinating multiple behavioral goals at various levels simultaneously (see also e.g., D'Esposito & Postle, 2015). This seems in line with highly influential multiple-component models of working memory, first introduced by Baddeley and Hitch in 1974. While temporal storages of these models allow buffering information (here: the goals on various levels of abstraction), the "central executive" is proposed to allow for manipulating and coordinating the information in a goal-directed manner (i.e., with regard to an overarching goal).

3.4. Limitations and Directions for Future Research

In future studies, it would be interesting to examine the relation between children's prospective processing of others' actions and their prospective action planning. Whereas the current thesis as well as previous research (e.g., Freier et al., 2015, 2017; Rosenbaum et al., 2012) hint at similar developmental pathways, it remains an open question whether a general capacity of processing hierarchical structures might account for children's improvements. In this context, it would be particularly interesting to investigate the role of more basic cognitive capacities, such as working memory, inhibitory control, or cognitive flexibility (commonly referred to as the three core executive functions, e.g., Miyake et al., 2000), as these basic capacities are supposed to be crucial to control behavior with respect to a desired action outcome and show significant improvements over early childhood (Diamond, 2013). Whereas the results of Study 3b point to an important role of inhibitory control and working memory for the prospective understanding of others' hierarchical actions, only a few studies have investigated the role of executive functions for children's prospective movement planning of hierarchical actions, and the findings are ambiguous (Scharoun, Gonzalez, et al., 2018; Stöckel & Hughes, 2016; Wunsch et al., 2016). Thus, taking a closer look at the similar developmental trajectories and possible underlying mechanisms might give us a broader picture of children's developing prospective action processing.

Furthermore, whereas Study 2 examined children's prospective action planning with respect to a specified action goal, it would be particularly interesting to investigate children's action planning within a social context. More precisely, one might ask whether young children prospectively plan their actions with respect to another person. This seems particularly interesting, as it would bring together children's prospective understanding of others' actions and children's prospective movement planning, which have hitherto been subject to rather dissociated areas of research (see Ansuini et al., 2018; Paulus, 2016 for exceptions).

Future research might furthermore investigate children's implicit prospective action understanding of others' complex actions. Given the theoretical considerations and findings of this thesis, one might expect that young children are not able to correctly visually anticipate actions that consist of several sub-actions (cf., complex actions in Study 3a and 3b). Instead, it might be hypothesized that they first learn to verbally predict such complex actions, before being able to visually anticipate them. To test this hypothesis, one might adopt the paradigm of Study 3a to an implicit version. Instead of explicitly stating the goal of the actor, children could be presented with an image depicting the final state of the action. Subsequently, their visual anticipations to either of the two paths (leading to the means or end) could be assessed.

Study 3b assessed German-speaking children to examine the development of hierarchical structure processing across the action and language domain. Whereas the structure of behavior seems at first glance comparable across cultures, languages vary in their syntactic structures (e.g., Haider, 2010). Interestingly, previous research indicated language-dependent differences in children's processing of grammatical structures (Kidd & Bavin, 2002; Lindner & Johnston, 1992). Therefore, it would be interesting to examine whether running the study with children of another language would reveal similar results. It could be speculated that if hierarchical structure plays a minor role in a certain language, children might develop the capacity of processing hierarchical language structures later in ontogeny (or not at all, see below) while showing a similar developmental pathway as German children for processing hierarchical actions. Relatedly, it would be particularly interesting to investigate the development of processing hierarchical structures in the Brazilian tribe known as the Pirahã. Daniel Everett who lived among the Pirahã for several years claims that their language has no embedding of phrases and is free of recursion (Everett, 2005). Whether the lack of hierarchical structure in the language domain is accompanied by a lack of processing hierarchical actions could broaden our understanding of hierarchical structure processing more generally.

Future research would furthermore profit from a conceptual clarification of hierarchy in the action domain. Although human behavior has commonly been referred to as hierarchical, a precise definition is missing. It is generally accepted that behavior consists of higher-level and lower-level elements (e.g., Botvinick, 2008; Byrne & Russon, 1998; Cooper et al., 2014; Miller et al., 1960; Pulvermüller & Fadiga, 2010), though, the nature of these elements and the relations between them are often inadequately specified and vary between approaches. This has also been pointed out by Uithol et al. (2012). His proposed alternative hierarchy of *temporal extension,* in which higher-level elements are represented longer and therefore influence an action over a longer period of time than lower-level elements, seems a promising approach. Nevertheless, it might turn out as a too general definition, especially for examining hierarchical structure processing in a developmental context. In any case, a more precise concept of action hierarchy should take the different kinds of proposed action hierarchies into account, not least to allow for empirical investigations of their relatedness. To give an example, Greenfield (1991) proposed that hierarchical organization in the first three years of life is reflected in children's strategy of nesting cups and in their increasing ability to combine words. However, it has been questioned whether these early competencies relate to children's later processing of more complex hierarchical structures (Karmiloff-Smith, 1997b). Taken together, a conceptual clarification seems the basis for a better understanding of hierarchical structure processing in early childhood.

Lastly, when investigating action processing in early childhood, one might question whether goals should be rather viewed as inscribed in the respective actions, as opposed to goals standing somewhat separate from the action, depicting an abstract entity that can be investigated on its own. In my view, young children prospectively process others' actions with respect to what they observe, that is, others' behavior. Within social interactions children then learn to talk and reason about others' actions, lifting their understanding of others' actions onto a new level. This new level might be considered as a prerequisite to be able to understand others' goals as mental states, abstracted from the action. Relatedly, although adults can flexibly reason about others' goals, desires, and beliefs and although they can use their knowledge about others' mental states to predict others' behavior, this reasoning might play a minor role in everyday prospective action processing (Apperly & Butterfill, 2009). It might be speculated that seeing someone reaching for the salt shaker and thereupon passing it over, has little to do with verbally reasoning about another's goal ("she wants the salt shaker, so I pass it over") but is rather an automatic reaction to the observed behavior.

3.5. Conclusion

This thesis provides new insights into how young children prospectively plan their actions and how they prospectively process others' actions of varying degrees of complexity.

First, it indicates that children's visual anticipations of others' simple grasping actions seem to be rather based on lower-level perceptual mechanisms, whereas no evidence could be found for motor simulation theories (Research Question 1). This questions interpretations of previous anticipation studies, which were taken as support for simulation theories, and highlights the role of lower-level perceptual mechanisms for children's visual anticipations.

Second, it points to an increase of controlled processes for children's prospective action planning as well as for their prospective processing of others' actions in early childhood (Research Question 2). It provides new empirical evidence that children might have to learn to reason about others' more complex behavior first (relying on controlled processes) before this knowledge is transferred into automatic visual anticipations (Research Question 2a). This conflicts with the predominant view that children understand others' actions implicitly before showing an explicit understanding later in ontogeny and opens new directions for developmental theorizing. Furthermore, the thesis provides new insights into children's own prospective action planning. Most interestingly, it indicates that prospective movement planning might develop as general capacity (Research Question 2b), expanding previous research, which has been inconclusive in this point.

Lastly, this thesis extends previous research on children's prospective understanding of other's simple actions to more complex actions, which seems important given the rather complex nature of everyday human behavior. It makes a novel contribution to the field of research by providing first direct empirical evidence that the processing of hierarchical action and language structures is related in early childhood and it suggests a more general role of working memory for hierarchical structure processing (Research Question 3).

4. References

Adalbjornsson, C. F., Fischman, M. G., & Rudisill, M. E. (2008). The end-state comfort effect in young children. *Research Quarterly for Exercise and Sport*, *79*(1), 36–41.

Adam, M., & Elsner, B. (2018). Action effects foster 11-month-olds' prediction of action goals for a non-human agent. *Infant Behavior & Development*, *53*, 49–55. https://doi.org/10.1016/j.infbeh.2018.09.002

Adam, M., Reitenbach, I., Papenmeier, F., Gredebäck, G., Elsner, C., & Elsner, B. (2016). Goal saliency boosts infants' action prediction for human manual actions, but not for mechanical claws. *Infant Behavior & Development*, *44*, 29–37. https://doi.org/10.1016/j.infbeh.2016.05.001

Aglioti, S. M., Cesari, P., Romani, M., & Urgesi, C. (2008). Action anticipation and motor resonance in elite basketball players. *Nature Neuroscience*, *11*(9), 1109–1116. https://doi.org/10.1038/nn.2182

Ambrosini, E., Costantini, M., & Sinigaglia, C. (2011). Grasping with the eyes. *Journal of Neurophysiology*, *106*(3), 1437–1442. https://doi.org/10.1152/jn.00118.2011

Ambrosini, E., Reddy, V., Looper, A. de, Costantini, M., Lopez, B., & Sinigaglia, C. (2013). Looking ahead: Anticipatory gaze and motor ability in infancy. *PloS ONE*, *8*(7), e67916. https://doi.org/10.1371/journal.pone.0067916

Ansuini, C., Podda, J., Battaglia, F. M., Veneselli, E., & Becchio, C. (2018). One hand, two hands, two people: Prospective sensorimotor control in children with autism.

Developmental Cognitive Neuroscience, *29*, 86–96. https://doi.org/10.1016/j.dcn.2017.02.009

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, *116*(4), 953–970. https://doi.org/10.1037/a0016923

Aslin, R. N. (2007). What's in a look? *Developmental Science*, *10*(1), 48–53. https://doi.org/10.1111/j.1467-7687.2007.00563.x

Baddeley, A. D., & Hitch, G. (1974). Working memory. In *Psychology of Learning and Motivation* (Vol. 8, pp. 47–89). Elsevier. https://doi.org/10.1016/S0079-7421(08)60452-1

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

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, *14*(3), 110–118. https://doi.org/10.1016/j.tics.2009.12.006

Baird, J. A., & Astington, J. W. (2005). Introduction: Why language matters. In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 2–25). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195159912.003.0001

Barone, P., Corradi, G., & Gomila, A. (2019). Infants' performance in spontaneous-response false belief tasks: A review and meta-analysis. *Infant Behavior & Development*, *57*, 101350. https://doi.org/10.1016/j.infbeh.2019.101350

Bello, A., Sparaci, L., Stefanini, S., Boria, S., Volterra, V., & Rizzolatti, G. (2014). A developmental study on children's capacity to ascribe goals and intentions to others. *Developmental Psychology*, *50*(2), 504–513. https://doi.org/10.1037/a0033375

Botvinick, M. M. (2008). Hierarchical models of behavior and prefrontal function. *Trends in Cognitive Sciences*, *12*(5), 201–208. https://doi.org/10.1016/j.tics.2008.02.009

Brandone, A. C., & Wellman, H. M. (2009). You can't always get what you want: Infants understand failed goal-directed actions. *Psychological Science*, *20*(1), 85–91. https://doi.org/10.1111/j.1467-9280.2008.02246.x

Byrne, R. W., & Russon, A. E. (1998). Learning by imitation: A hierarchical approach. *Behavioral and Brain Sciences*, *21*(5), 667-684. https://doi.org/10.1017/s0140525x98001745

Cannon, E. N., & Woodward, A. L. (2012). Infants generate goal-based action predictions. *Developmental Science*, *15*(2), 292–298. https://doi.org/10.1111/j.1467-7687.2011.01127.x

Cannon, E. N., Woodward, A. L., Gredebäck, G., Hofsten, C. von, & Turek, C. (2012). Action production influences 12-month-old infants' attention to others' actions. *Developmental Science*, *15*(1), 35–42. https://doi.org/10.1111/j.1467-7687.2011.01095.x.

Carpendale, J. I. M., & Lewis, C. (2004). Constructing an understanding of mind: The development of children's social understanding within social interaction. *Behavioral and Brain Sciences*, *27*(1), 79-96. https://doi.org/10.1017/s0140525x04000032

Chomsky, N. (1956). Three models for the description of language. *IRE Transactions on Information Theory*, *2*(3), 113–124. https://doi.org/10.1109/TIT.1956.1056813

Cooper, R. P., Ruh, N., & Mareschal, D. (2014). The goal circuit model: A hierarchical multi-route model of the acquisition and control of routine sequential action in humans. *Cognitive Science*, *38*(2), 244–274. https://doi.org/10.1111/cogs.12067

Corrêa, L. M. S. (1995). An alternative assessment of children's comprehension of relative clauses. *Journal of Psycholinguistic Research*, *24*(3), 183–203. https://doi.org/10.1007/BF02145355

Daum, M. M., Attig, M., Gunawan, R., Prinz, W., & Gredebäck, G. (2012). Actions seen through babies' eyes: A dissociation between looking time and predictive gaze. *Frontiers in Psychology*, *3*, 370. https://doi.org/10.3389/fpsyg.2012.00370

Daum, M. M., Prinz, W., & Aschersleben, G. (2011). Perception and production of object-related grasping in 6-month-olds. *Journal of Experimental Child Psychology*, *108*(4), 810–818. https://doi.org/10.1016/j.jecp.2010.10.003

Daum, M. M., Vuori, M. T., Prinz, W., & Aschersleben, G. (2009). Inferring the size of a goal object from an actor's grasping movement in 6- and 9-month-old infants. *Developmental Science*, *12*(6), 854–862. https://doi.org/10.1111/j.1467-7687.2009.00831.x

Dawkins, R. (1976). Hierarchical organisation: A candidate principle for ethology. In P. P. G. Bateson & R. A. Hinde (Eds.), *Growing points in ethology.* Cambridge University Press.

de Wit, S., & Dickinson, A. (2009). Associative theories of goal-directed behaviour: A case for animal-human translational models. *Psychological Research*, *73*(4), 463–476. https://doi.org/10.1007/s00426-009-0230-6

de Wit, S., Watson, P., Harsay, H. A., Cohen, M. X., van de Vijver, I., & Ridderinkhof, K. R. (2012). Corticostriatal connectivity underlies individual differences in the balance between habitual and goal-directed action control. *Journal of Neuroscience*, *32*(35), 12066–12075. https://doi.org/10.1523/JNEUROSCI.1088-12.2012

DeKeyser, R. (2015). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (2nd ed.). Routledge.

Delabarre, E. B. (1898). A method of recording eye-movements. *The American Journal of Psychology*, *9*(4), 572. https://doi.org/10.2307/1412191

D'Esposito, M., & Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annual Review of Psychology*, *66*, 115–142. https://doi.org/10.1146/annurev-psych-010814-015031

Devine, R. T., & Hughes, C. (2014). Relations between false belief understanding and executive function in early childhood: A meta-analysis. *Child Development*, *85*(5), 1777–1794. https://doi.org/10.1111/cdev.12237

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: A neurophysiological study. *Experimental Brain Research*, *91*(1), 176–180. https://doi.org/10.1007/BF00230027

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Elsner, B., & Adam, M. (2020). Infants' goal prediction for simple action events: The role of experience and agency cues. *Topics in Cognitive Science*, *13*(1), 45–62. https://doi.org/10.1111/tops.12494

Everett, D. (2005). Cultural constraints on grammar and cognition in Pirahã: Another look at the design features of human language. *Current Anthropology*, *46*(4), 621–646. https://doi.org/10.1086/431525

Fadiga, L., Craighero, L., & D'Ausilio, A. (2009). Broca's area in language, action, and music. *Annals of the New York Academy of Sciences*, *1169*, 448–458. https://doi.org/10.1111/j.1749-6632.2009.04582.x

Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: A magnetic stimulation study. *Journal of Neurophysiology*, *73*(6), 2608–2611. https://doi.org/10.1152/jn.1995.73.6.2608

Falck-Ytter, T., Gredebäck, G., & Hofsten, C. von (2006). Infants predict other people's action goals. *Nature Neuroscience*, *9*(7), 878–879. https://doi.org/10.1038/nn1729

Fantz, R. L. (1958). Pattern vision in young infants. *The Psychological Record*, *8*(2), 43–47. https://doi.org/10.1007/BF03393306

Fengler, A., Meyer, L., & Friederici, A. D. (2015). Brain structural correlates of complex sentence comprehension in children. *Developmental Cognitive Neuroscience*, *15*, 48–57. https://doi.org/10.1016/j.dcn.2015.09.004

Fengler, A., Meyer, L., & Friederici, A. D. (2016). How the brain attunes to sentence processing: Relating behavior, structure, and function. *NeuroImage*, *129*, 268–278. https://doi.org/10.1016/j.neuroimage.2016.01.012

Flanagan, J. R., & Johansson, R. S. (2003). Action plans used in action observation. *Nature*, *424*, 769–771. https://doi.org/10.1038/nature01861

Flynn, E., & Whiten, A. (2008). Imitation of hierarchical structure versus component details of complex actions by 3- and 5-year olds. *Journal of Experimental Child Psychology*, *101*(4), 228–240. https://doi.org/10.1016/.jecp.2008.05.009

Freier, L., Cooper, R. P., & Mareschal, D. (2015). The planning and execution of natural sequential actions in the preschool years. *Cognition*, *144*, 58–66. https://doi.org/10.1016/j.cognition.2015.07.005

Freier, L., Cooper, R. P., & Mareschal, D. (2017). Preschool children's control of action outcomes. *Developmental Science*, *20*(2), e12354. https://doi.org/10.1111/desc.12354

Friederici, A. D., Bahlmann, J., Friedrich, R., & Makuuchi, M. (2011). The neural basis of recursion and complex syntactic hierarchy. *Biolinguistics*, *5*(1-2), 87–104.

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development*, *10*(4), 483–527. https://doi.org/10.1016/0885-2014(95)90024-1

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119 (Pt 2)*, 593–609. https://doi.org/10.1093/brain/119.2.593

Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences*, *8*(9), 396–403. https://doi.org/10.1016/j.tics.2004.07.002

Ganglmayer, K., Attig, M., Daum, M. M., & Paulus, M. (2019). Infants' perception of goal-directed actions: A multi-lab replication reveals that infants anticipate paths and not goals. *Infant Behavior & Development*, *57*, 101340. https://doi.org/10.1016/j.infbeh.2019.101340

Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, *56*(2), 165–193. https://doi.org/10.1016/0010-0277(95)00661-H

Goldman, A. I. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading*. *Philosophy of mind*. Oxford University Press.

Gordon, R. M. (1986). Folk psychology as simulation. *Mind & Language*, *1*(2), 158–171. https://doi.org/10.1111/j.1468-0017.1986.tb00324.x

Gredebäck, G., & Daum, M. M. (2015). The microstructure of action perception in infancy: Decomposing the temporal structure of social information processing. *Child Development Perspectives*, *9*(2), 79–83. https://doi.org/10.1111/cdep.12109

Greenfield, P. M. (1991). Language, tools and brain: The ontogeny and phylogeny of hierarchically organized sequential behavior. *Behavioral and Brain Sciences*, *14*(04), 531–595. https://doi.org/10.1017/S0140525X00071235

Greenfield, P. M., & Schneider, L. (1977). Building a tree structure: The development of hierarchical complexity and interrupted strategies in children's construction activity.

Developmental Psychology, *13*(4), 299–313. https://doi.org/10.1037/0012-1649.13.4.299

Grosse Wiesmann, C., Friederici, A. D., Singer, T., & Steinbeis, N. (2017). Implicit and explicit false belief development in preschool children. *Developmental Science*, *20*(5), e12445. https://doi.org/10.1111/desc.12445

Grossman, M. (1980). A central processor for hierarchically-structured material: Evidence from Broca's aphasia. *Neuropsychologia*, *18*(3), 299–308. https://doi.org/10.1016/0028-3932(80)90125-6

Haider, H. (2010). *The syntax of German. Cambridge syntax guides*. Cambridge University Press. https://doi.org/10.1017/CBO9780511845314

Haith, M. M. (1998). Who put the cog in infant cognition? Is rich interpretation too costly? *Infant Behavior and Development*, *21*(2), 167–179. https://doi.org/10.1016/S0163-6383(98)90001-7

Henrichs, I., Elsner, C., Elsner, B., & Gredebäck, G. (2012). Goal salience affects infants' goal-directed gaze shifts. *Frontiers in Psychology*, *3*, 391. https://doi.org/10.3389/fpsyg.2012.00391

Heyes, C. (2014). False belief in infancy: A fresh look. *Developmental Science*, *17*(5), 647–659. https://doi.org/10.1111/desc.12148

Heyes, C., & Catmur, C. (2020). What happened to mirror neurons? *Perspectives on Psychological Science.* Advance online publication. https://doi.org/10.31234/osf.io/dtnqg

Hickok, G. (2014). *The myth of mirror neurons: The real neuroscience of communication and cognition* (1. ed.). Norton.

Hofmann, W., Friese, M., & Strack, F. (2009). Impulse and self-control from a dual-systems perspective. *Perspectives on Psychological Science*, *4*(2), 162–176. https://doi.org/10.1111/j.1745-6924.2009.01116.x

Hofsten, C. von, Kochukhova, O., & Rosander, K. (2007). Predictive tracking over occlusions by 4-month-old infants. *Developmental Science*, *10*(5), 625–640. https://doi.org/10.1111/j.1467-7687.2007.00604.x

Hunnius, S., & Bekkering, H. (2010). The early development of object knowledge: A study of infants' visual anticipations during action observation. *Developmental Psychology*, *46*(2), 446–454. https://doi.org/10.1037/a0016543

Hunnius, S., & Bekkering, H. (2014). What are you doing? How active and observational experience shape infants' action understanding. *Philosophical Transactions of the Royal Society B*, *369*(1644), 20130490. https://doi.org/10.1098/rstb.2013.0490

Jacob, P. (2009). The tuning-fork model of human social cognition: A critique. *Consciousness and Cognition*, *18*(1), 229–243. https://doi.org/10.1016/j.concog.2008.05.002

Jeon, H. A. (2014). Hierarchical processing in the prefrontal cortex in a variety of cognitive domains. *Frontiers in Systems Neuroscience*, *8*, 223. https://doi.org/10.3389/fnsys.2014.00223

Jovanovic, B., & Schwarzer, G. (2017). The influence of grasping habits and object orientation on motor planning in children and adults. *Developmental Psychobiology*, *59*(8), 949–957. https://doi.org/10.1002/dev.21573

Kanakogi, Y., & Itakura, S. (2011). Developmental correspondence between action prediction and motor ability in early infancy. *Nature Communications*, *2*, 341. https://doi.org/10.1038/ncomms1342

Karmiloff-Smith, A. (1997a). *Beyond modularity: A developmental perspective on cognitive science* (3rd printing). *Learning, development, and conceptual change*. MIT Press.

Karmiloff-Smith, A. (1997b). Précis of beyond modularity: A developmental perspective on cognitive science. *Behavioral and Brain Sciences*, *17*(4), 693–707. https://doi.org/10.1017/S0140525X00036621

Kenward, B., Folke, S., Holmberg, J., Johansson, A., & Gredebäck, G. (2009). Goal directedness and decision making in infants. *Developmental Psychology*, *45*(3), 809–819. https://doi.org/10.1037/a0014076

Kidd, E., & Bavin, E. L. (2002). English-speaking children's comprehension of relative clauses: evidence for general-cognitive and language-specific constraints on development. *Journal of Psycholinguistic Research*, *31*(6), 599–617. https://doi.org/10.1023/A:1021265021141

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35-B42. https://doi.org/10.1016/S0010-0277(02)00004-5

Kirkham, N. Z., Slemmer, J. A., Richardson, D. C., & Johnson, S. P. (2007). Location, location, location: Development of spatiotemporal sequence learning in infancy. *Child Development*, *78*(5), 1559–1571. https://doi.org/10.1111/j.1467-8624.2007.01083.x

Klossek, U. M. H., Russell, J., & Dickinson, A. (2008). The control of instrumental action following outcome devaluation in young children aged between 1 and 4 years. *Journal of Experimental Psychology: General*, *137*(1), 39–51. https://doi.org/10.1037/0096-3445.137.1.39

Knudsen, B., Henning, A., Wunsch, K., Weigelt, M., & Aschersleben, G. (2012). The end-state comfort effect in 3- to 8-year-old children in two object manipulation tasks. *Frontiers in Psychology*, *3*, 445. https://doi.org/10.3389/fpsyg.2012.00445

Kochukhova, O., & Gredebäck, G. (2010). Preverbal infants anticipate that food will be brought to the mouth: An eye tracking study of manual feeding and flying spoons. *Child Development*, *81*(6), 1729–1738. https://doi.org/10.1111/j.1467-8624.2010.01506.x

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jefress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–136). New York: Wiley.

Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition*, *25*(3), 265–288. https://doi.org/10.1016/S0010-0277(87)80006-9

Lieberman, M. D. (2007). Social cognitive neuroscience: A review of core processes. *Annual Review of Psychology*, *58*, 259–289. https://doi.org/10.1146/annurev.psych.58.110405.085654

Lindner, K., & Johnston, J. R. (1992). Grammatical morphology in language–impaired children acquiring English or German as their first language: A functional perspective. *Applied Psycholinguistics*, *13*(2), 115–129.

Low, J., & Perner, J. (2012). Implicit and explicit theory of mind: State of the art. *British Journal of Developmental Psychology*, *30*(1), 1–13. https://doi.org/10.1111/j.2044-835X.2011.02074.x

Marcus, G. F. (2006). Cognitive architecture and descent with modification. *Cognition*, *101*(2), 443–465. https://doi.org/10.1016/j.cognition.2006.04.009

Martins, M. J. D., Bianco, R., Sammler, D., & Villringer, A. (2019). Recursion in action: An fMRI study on the generation of new hierarchical levels in motor sequences. *Human Brain Mapping*, *40*(9), 2623–2638. https://doi.org/10.1002/hbm.24549

Meyer, L., Obleser, J., & Friederici, A. D. (2013). Left parietal alpha enhancement during working memory-intensive sentence processing. *Cortex*, *49*(3), 711–721. https://doi.org/10.1016/j.cortex.2012.03.006

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior*. Henry Holt and Co. https://doi.org/10.1037/10039-000

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100.

Monroy, C. D., Gerson, S. A., & Hunnius, S. (2017). Toddlers' action prediction: Statistical learning of continuous action sequences. *Journal of Experimental Child Psychology*, *157*, 14–28. https://doi.org/10.1016/j.jecp.2016.12.004

Monroy, C. D., Meyer, M., Gerson, S., & Hunnius, S. (2017). Statistical learning in social action contexts. *PloS ONE*, *12*(5), e0177261. https://doi.org/10.1371/journal.pone.0177261

Pastra, K., & Aloimonos, Y. (2012). The minimalist grammar of action. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1585), 103–117. https://doi.org/10.1098/rstb.2011.0123

Paulus, M. (2012). Action mirroring and action understanding: An ideomotor and attentional account. *Psychological Research*, *76*(6), 760–767. https://doi.org/10.1007/s00426-011-0385-9

Paulus, M. (2016). The development of action planning in a joint action context. *Developmental Psychology*, *52*(7), 1052–1063. https://doi.org/10.1037/dev0000139

Paulus, M. (2021). *Should infant psychology rely on the violation-of-expectation method? Not anymore* [Manuscript submitted for publication].

Paulus, M., Hunnius, S., van Elk, M., & Bekkering, H. (2012). How learning to shake a rattle affects 8-month-old infants' perception of the rattle's sound: Electrophysiological evidence for action-effect binding in infancy. *Developmental Cognitive Neuroscience*, *2*(1), 90–96. https://doi.org/10.1016/j.dcn.2011.05.006

Paulus, M., Schuwerk, T., Sodian, B., & Ganglmayer, K. (2017). Children's and adults' use of verbal information to visually anticipate others' actions: A study on explicit and implicit social-cognitive processing. *Cognition*, *160*, 145–152. https://doi.org/10.1016/j.cognition.2016.12.013

Pereira, K., Medeiros, J. C., Bernardes, L. R., & Souza, L. A. P. S. de (2019). End-state comfort effect in manipulative motor actions of typical and atypical children: a systematic review. *Fisioterapia E Pesquisa*, *26*(2), 202–209. https://doi.org/10.1590/1809-2950/17009426022019

Pomiechowska, B., & Csibra, G. (2017). Motor activation during action perception depends on action interpretation. *Neuropsychologia*, *105*, 84–91. https://doi.org/10.1016/j.neuropsychologia.2017.01.032

Pulvermüller, F. (2014). The syntax of action. *Trends in Cognitive Sciences*, *18*(5), 219–220. https://doi.org/10.1016/j.tics.2014.01.001

Pulvermüller, F., & Fadiga, L. (2010). Active perception: Sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, *11*(5), 351–360. https://doi.org/10.1038/nrn2811

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews. Neuroscience*, *2*(9), 661–670. https://doi.org/10.1038/35090060

Rizzolatti, G., & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: Interpretations and misinterpretations. *Nature Reviews. Neuroscience*, *11*(4), 264–274. https://doi.org/10.1038/nrn2805

Rosander, K., & Hofsten, C. von (2011). Predictive gaze shifts elicited during observed and performed actions in 10-month-old infants and adults. *Neuropsychologia*, *49*(10), 2911–2917. https://doi.org/10.1016/j.neuropsychologia.2011.06.018

Rosenbaum, D. A. (2009). *Human motor control* (2nd ed.). Academic Press: Elsevier Inc.

Rosenbaum, D. A., Chapman, K. M., Weigelt, M., Weiss, D. J., & van der Wel, R. (2012). Cognition, action, and object manipulation. *Psychological Bulletin*, *138*(5), 924–946. https://doi.org/10.1037/a0027839

Ruffman, T. (2014). To belief or not belief: Children's theory of mind. *Developmental Review*, *34*(3), 265–293. https://doi.org/10.1016/j.dr.2014.04.001

Ruffman, T., Taumoepeau, M., & Perkins, C. (2012). Statistical learning as a basis for social understanding in children. *British Journal of Developmental Psychology*, *30*(1), 87–104. https://doi.org/10.1111/j.2044-835X.2011.02045.x

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science (New York, N.Y.)*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Saffran, J. R., & Kirkham, N. Z. (2018). Infant statistical learning. *Annual Review of Psychology*, *69*, 181–203. https://doi.org/10.1146/annurev-psych-122216-011805

Sartori, L., Becchio, C., & Castiello, U. (2011). Cues to intention: The role of movement information. *Cognition*, *119*(2), 242–252. https://doi.org/10.1016/j.cognition.2011.01.014

Scharoun, S. M., Gonzalez, D. A., Roy, E. A., & Bryden, P. J. (2018). End-State comfort across the lifespan: A cross-sectional investigation of how movement context influences motor planning in an overturned glass task. *Motor Control*, *22*(2), 211–230. https://doi.org/10.1123/mc.2016-0064

Scharoun, S. M., Robinson, L. E., Logan, S. W., Bryden, P. J., Brock, S. J., & Fischman, M. G. (2018). Sex differences in the end-state comfort effect in pre-adolescent children. *Human Movement Science*, *57*, 244–250. https://doi.org/10.1016/j.humov.2017.09.001

Scharoun Benson, S. M., Roy, E. A., & Bryden, P. J. (2018). End-state comfort in two object manipulation tasks: Investigating how the movement context influences planning in children, young adults, and older adults. *Developmental Psychobiology*, *60*(3), 317–323. https://doi.org/10.1002/dev.21611

Schuwerk, T., & Paulus, M. (2016). Preschoolers, adolescents, and adults visually anticipate an agent's efficient action; but only after having observed it frequently. *Quarterly Journal of Experimental Psychology*, *69*(4), 800–816. https://doi.org/10.1080/17470218.2015.1061028

Sebanz, N., & Knoblich, G. (2009). Prediction in joint action: What, when, and where. *Topics in Cognitive Science*, *1*(2), 353–367. https://doi.org/10.1111/j.1756-8765.2009.01024.x

Siegler, R. S., Eisenberg, N., DeLoache, J. S., & Saffran, J. (2016). *Entwicklungspsychologie im Kindes- und Jugendalter* (4. Auflage). Springer. https://doi.org/10.1007/978-3-662-47028-2

Sodian, B. (2016). Is false belief understanding continuous from infancy to preschool age? In D. Barner & A. S. Baron (Eds.), *Core knowledge and conceptual change* (pp. 301–320). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190467630.003.0017

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

Sommerville, J. A., Hildebrand, E. A., & Crane, C. C. (2008). Experience matters: The impact of doing versus watching on infants' subsequent perception of tool-use events. *Developmental Psychology*, *44*(5), 1249–1256. https://doi.org/10.1037/a0012296

Sommerville, J. A., Woodward, A. L., & Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. *Cognition*, *96*(1), B1-11. https://doi.org/10.1016/j.cognition.2004.07.004

Southgate, V. (2013). Do infants provide evidence that the mirror system is involved in action understanding? *Consciousness and Cognition*, *22*(3), 1114–1121. https://doi.org/10.1016/j.concog.2013.04.008

Southgate, V., Johnson, M. H., Osborne, T., & Csibra, G. (2009). Predictive motor activation during action observation in human infants. *Biology Letters*, *5*(6), 769–772. https://doi.org/10.1098/rsbl.2009.0474

Stöckel, T., & Hughes, C. M. L. (2016). The relation between measures of cognitive and motor functioning in 5- to 6-year-old children. *Psychological Research*, *80*(4), 543–554. https://doi.org/10.1007/s00426-015-0662-0

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, *8*(3), 220–247. https://doi.org/10.1207/s15327957pspr0803_1

Theeuwes, J., & Belopolsky, A. (2010). Top–down and bottom–up control of visual selection: controversies and debate. In V. Coltheart (Ed.), *Tutorials in Visual Cognition.* Psychology Press.

Thibaut, J.-P., & Toussaint, L. (2010). Developing motor planning over ages. *Journal of Experimental Child Psychology*, *105*(1-2), 116–129. https://doi.org/10.1016/j.jecp.2009.10.003

Thompson, E. L., Bird, G., & Catmur, C. (2019). Conceptualizing and testing action understanding. *Neuroscience & Biobehavioral Reviews*, *105*, 106–114. https://doi.org/10.1016/j.neubiorev.2019.08.002

Tomasello, M. (2014). The ultra-social animal. *European Journal of Social Psychology*, *44*(3), 187–194. https://doi.org/10.1002/ejsp.2015

Uddén, J., Martins, M. J. D. de, Zuidema, W., & Tecumseh Fitch, W. (2020). Hierarchical structure in sequence processing: How to measure it and determine its neural implementation. *Topics in Cognitive Science*, *12*(3), 910–924. https://doi.org/10.1111/tops.12442

Uithol, S., & Paulus, M. (2014). What do infants understand of others' action? A theoretical account of early social cognition. *Psychological Research*, *78*(5), 609–622. https://doi.org/10.1007/s00426-013-0519-3

Uithol, S., van Rooij, I., Bekkering, H., & Haselager, P. (2012). Hierarchies in action and motor control. *Journal of Cognitive Neuroscience*, *24*(5), 1077–1086.

Villiers, J. G. de, Flusberg, H. B. T., Hakuta, K., & Cohen, M. (1979). Children's comprehension of relative clauses. *Journal of Psycholinguistic Research*, *8*(5), 499–518. https://doi.org/10.1007/BF01067332

Weigelt, M., & Schack, T. (2010). The development of end-state comfort planning in preschool children. *Experimental Psychology*, *57*(6), 476–782. https://doi.org/10.1027/1618-3169/a000059

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

Whiten, A., Flynn, E., Brown, K., & Lee, T. (2006). Imitation of hierarchical action structure by young children. *Developmental Science*, *9*(6), 574–582. https://doi.org/10.1111/j.1467-7687.2006.00535.x

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69*(1), 1–34. https://doi.org/10.1016/S0010-0277(98)00058-4

Woodward, A. L., & Gerson, S. A. (2014). Mirroring and the development of action understanding. *Philosophical Transactions of the Royal Society B*, *369*(1644), 20130181. https://doi.org/10.1098/rstb.2013.0181

Woodward, A. L., Sommerville, J. A., Gerson, S., Henderson, A. M., & Buresh, J. (2009). The emergence of intention attribution in infancy. In *Psychology of Learning and Motivation. The Psychology of Learning and Motivation* (Vol. 51, pp. 187–222). Elsevier. https://doi.org/10.1016/S0079-7421(09)51006-7

Wunsch, K., Henning, A., Aschersleben, G., & Weigelt, M. (2013). A systematic review of the end-state comfort effect in normally developing children and in children with developmental disorders. *Journal of Motor Learning and Development*, *1*(3), 59–76. https://doi.org/10.1123/jmld.1.3.59

Wunsch, K., Pfister, R., Henning, A., Aschersleben, G., & Weigelt, M. (2016). No interrelation of motor planning and executive functions across young ages. *Frontiers in Psychology*, *7*, 1031. https://doi.org/10.3389/fpsyg.2016.01031

Wunsch, K., & Weigelt, M. (2016). A three-stage model for the acquisition of anticipatory planning skills for grip selection during object manipulation in young children. *Frontiers in Psychology*, *7*, 958. https://doi.org/10.3389/fpsyg.2016.00958

Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, *358*(6389), 749–750. https://doi.org/10.1038/358749a0

Yarbus, A. L. (1967). *Eye movements and vision*. Plenum Press.

Zaccarella, E., Papitto, G., & Friederici, A. D. (2021). Language and action in Broca's area: Computational differentiation and cortical segregation. *Brain and Cognition*, *147*, 105651. https://doi.org/10.1016/j.bandc.2020.105651

5. Appendices

A) Study 1: The development of children’s use of kinematic cues for action anticipation and action prediction

Melzel, S., Altvater-Mackensen, N., Ganglmayer, K., Müller, F., Hauf, P., & Paulus, M. (submitted). The development of children’s use of kinematic cues for action anticipation and action prediction.

Abstract

Expectations about how others' actions unfold in the future are crucial for our everyday social interactions. The current study examined the development of the use of kinematic cues for action anticipation and prediction in 3-year-olds, 4-year-olds, 10-year-olds and adults (*n* = 178) in two experiments. Participants observed a hand repeatedly reaching for either a close or a far object. The motor kinematics of the hand varied depending on whether the hand reached for the close or far object, respectively. We assessed whether participants would use kinematic cues to visually anticipate (Experiment 1) and verbally predict (Experiment 2) which object the hand was going to grasp. We found that only adults but not 3- to10-year-olds based their visual anticipations on kinematic cues (Experiment 1). This speaks against claims that action anticipations are based on simulating others' motor processes and instead provides evidence that anticipations are based on perceptual mechanisms. Interestingly, 10-year-olds used kinematic cues to correctly verbally predict the target-object and 4-year-olds learned to do so over the trials (Experiment 2). Thus, kinematic cues are used earlier in life for explicit action predictions than for visual action anticipations. This adds to a recent debate on whether or not an implicit understanding of others' actions precedes their ability to verbally reason about the same actions.

Keywords: action anticipation, action prediction, simulation theory, low level mechanisms, implicit, explicit

The development of children's use of kinematic cues for action anticipation and action prediction

Understanding others and having an expectation about how their action unfolds in the future is crucial for our everyday social interactions (Sebanz & Knoblich, 2009). It allows for cooperation with others (Bekkering et al., 2009), fluent interactions (Meyer et al., 2016), and successful task completion (Brownell & Carriger, 1990). Thus, investigating its development has become a key question in developmental science (Bartsch & Wellman, 1989; Carpendale & Lewis, 2004; Ganglmayer, Attig, et al., 2019; Kayhan et al., 2019; Monroy et al., 2018).

How do children learn to predict others' behavior? One theoretical framework highlights the role of statistical learning and the role of perceptual cues in anticipating and predicting others' actions (e.g., Hunnius & Bekkering, 2014; Ruffman et al., 2012; see also Saffran & Kirkham, 2018; Smith et al., 2018). Indeed, first empirical studies demonstrated that infants predict an upcoming action based on how this action was performed previously (e.g., Paulus et al., 2011). Furthermore, empirical work with adults suggests that they rely on a nearest-object-heuristic: adults anticipate the nearest out of several objects being placed in the direction of a movement when they have no further information about the target of the action (Rotman et al., 2006).

Another framework stresses the notion of internal simulation processes. According to simulation theories we understand and predict other's actions by using our own mind to simulate the mental processes of others (e.g., Gallese et al., 2004; Goldman, 2006; Gordon, 1995; Rizzolatti et al., 2001). One prominent simulation approach is the direct-matching hypothesis of action understanding according to which we understand others' action goals by matching an observed action onto our own motor repertoire (Gallese et al., 2004; Rizzolatti et al., 2001; Rizzolatti & Sinigaglia, 2010). As pro-active gaze movements have been found to be crucial to perform visually guided actions, the presence of similar anticipatory eye movements when observing someone else performing the same action have been taken as evidence for the direct-matching account (Falck-Ytter et al., 2006; Flanagan & Johansson, 2003; Kanakogi & Itakura, 2011).

Notably, to investigate the early development of children's expectations about others' actions, different types of measures have been used: Non-verbal gaze measures such as looking time or visual anticipations (often referred to as *implicit* measures) and verbal predictions (often referred to as *explicit* measures). Anticipation studies have shown that children start to perceive others' actions as target-directed during the first year of life—long before they are able to verbally predict others' actions (e.g., Cannon & Woodward, 2012; Falck-Ytter et al., 2006; Hunnius & Bekkering, 2010; Kochukhova & Gredebäck, 2010). This line of research has led to the predominant view that children develop an implicit understanding of others' actions before showing an explicit understanding. This "implicit precedes explicit understanding account" suggests that young children first implicitly understand others' actions before they are able to translate their implicit understanding in explicit terms, that is, into language (e.g., Baillargeon et al., 2010). However, recent studies indicate that explicitly predicting another's action sometimes precedes correctly anticipating the same action (e.g., Paulus et al., 2017). According to an "explicit proceeds implicit understanding account" more complex forms of understanding are first acquired on a verbal and explicit level. With increasing experience within this area and automatization, this knowledge can then be transformed into an implicit way of understanding. This is similar to a situation in which one learns to drive a car: after explicit instruction the knowledge and routines become automatized. Overall, these views lead to the interesting developmental question to which extend an implicit or explicit understanding of others' actions develop earlier (Barone et al., 2019; Grosse Wiesmann et al., 2017; Paulus et al., 2017). Our study aimed at contributing to this debate.

Overall, the current study has two main aims: First, to contribute to the debate whether children's action anticipation is based on motor simulation of others' actions or whether action anticipations are rather based on lower level mechanisms. Second, to examine whether children show successful visual action anticipation and verbal action prediction at the same age or whether one precedes the other, adding to discussions about the developmental sequence of implicit and explicit action understanding. In the following, we introduce the central theoretical views that are in the focus of the current study.

Different aspects of action understanding: visual anticipations and verbal predictions

Action understanding has been used as an umbrella term and comprises various behaviors and competencies. It comprises both non-verbal aspects as well as verbal aspects (often referred to as *implicit* and *explicit* understanding; e.g., Low & Perner, 2012). A common measure to assess implicit action understanding are visual anticipations (for review see Hunnius & Bekkering, 2014). Explicit action understanding implies processing verbal information and has frequently been assessed by verbal predictions (e.g., Wellman & Woolley, 1990).

Previous findings suggest that infants start to perceive others' actions as target-directed during the first year of life (e.g., Cannon & Woodward, 2012; Falck-Ytter et al., 2006; Hunnius & Bekkering, 2010; Kochukhova & Gredebäck, 2010). For example, Falck-Ytter et al. (2006) showed that 12-month-olds anticipate a bucket in which an actor is about to place a toy and Hunnius and Bekkering (2010) found that already 6-month-olds anticipate to the mouth region of an actor when observing the actor lifting a cup. Furthermore, it has been claimed that already infants base their anticipations on mental states of others (such as goals, beliefs and desires) and that they master implicit false-belief tasks during the second year of life (Onishi & Baillargeon, 2005; Southgate et al., 2007), whereas they explicitly master such tasks at around the age of four (Wellman et al., 2001). Given the limited language abilities during the first two years of life this line of research has led to the predominant view that children show an implicit understanding of others' actions before they are able to verbally reason about the same actions.

However, there is a recent debate whether this is always the case (Barone et al., 2019; Grosse Wiesmann et al., 2017). Indeed, other studies have shown that children first correctly verbally predict others' actions before correctly anticipating them (Paulus et al., 2017; Schuwerk & Paulus, 2016). For example, Paulus et al. (2017) showed that 2.5-year-olds use verbal information about an actor's goal to predict the action but did not use the information to correctly anticipate the action. Thus, children might first learn to reason about others' behavior before correctly anticipating it. In a similar way, theories on skill acquisition suggest that sometimes people first need to learn a novel behavior (e.g., driving a car) through explicit training before it becomes an (implicit) automatized skill. In this sense, correctly anticipating others' behavior might constitute a skill that capitalizes on rich experiences of reasoning and talking about others' actions (cf. Carpendale & Lewis, 2004).

Action anticipation: simulation theory vs. lower level mechanisms

According to simulation theories we understand and predict others' actions by simulating others' mental processes (e.g., Gallese et al., 2004; Goldman, 2006; Gordon, 1995; Paulus, 2012; Rizzolatti et al., 2001). Whereas simulation theories differ in the presumed cognitive levels involved in the simulation process (e.g., simulation of the movement vs. the goal or intention of the actor) they agree that simulating an observed action leads to an expectation about how the action unfolds. One prominent simulation approach is the direct-matching hypothesis. According to this approach observing an action elicits a motor activation in the observer's cognitive system which is similar to the one the observer has when performing the action themselves and which allows to "understand directly the goal of the actions of others without needing inferential processing" (Rizzolatti & Sinigaglia, 2010, p. 268). Indeed, there is evidence indicating a close relation between action production and action perception in children (Ambrosini et al., 2013; Daum et al., 2011; Kanakogi & Itakura, 2011; Kochukhova & Gredebäck, 2010; Rosander & Hofsten, 2011; Southgate et al., 2009). For example, Falck-Ytter et al. (2006) showed that 12-months-olds and adults but not 6-month-olds visually anticipate to a bucket in which an actor is about to place a toy. They interpreted their findings as support of the direct-matching hypothesis: only 12-month-olds and adults but not 6-months-olds master the observed action themselves, can thus match the observed action onto their own motor repertoire, and are therefore able to anticipate the target of the action (i.e., the bucket). However, they used a single, highly salient target-object (a red bucket with a 3D happy face connected to it) that has been placed in the direction of the placing movement. Thus, it cannot be ruled out that anticipations were based on lower level perceptual mechanisms. There seem to be at least two alternative explanations, which will be discussed in the following.

Noteworthy, spatial relationships seem to play a key role during learning to anticipate repetitive events in the visual domain (Saffran & Kirkham, 2018). Associations formed between an action and the location at which the action is typically directed at as well as prototypical movement patterns learned by repeated observations might be sufficient to elicit anticipatory looking. With respect to the results of Falck-Ytter et al. (2006) 12-month-olds have observed considerably more placing actions than 6-month-olds. By extracting regularities across several observations they might have learned that a placement movement usually continues in its initial direction. Thus, the results might also be explained by 12-month-olds opposed to 6-month-olds being able to anticipate the general movement direction until their gaze hits an interesting object located within that direction. Interestingly, a computational model resting on the assumption that the recognition of biological movements is based on learned prototypical patterns has been proposed to account for many movement recognition results (Giese & Poggio, 2003). The authors argue that “attention and top-down influences are not necessary” for basic motion recognition (Giese & Poggio, 2003, p. 190). Such learned prototypical motion patterns might also account for anticipatory eye movements.

Furthermore, gaze shifts can be triggered by non-foveal retinal stimulation (Harris, 1989). That is, the probability that a saccade is triggered by a certain area of a stimulus depends on the saliency of this area. The salience of a certain area is specified by the visual features of that area, and, due to the spatial inhomogeneity of the retina, by its locus on the retina. To give an example: imagine you are looking at a field of daisies. Now imagine a red tulip on the left and a white tulip on the right both equally distanced from your current fixation. Due to the higher saliency your first saccade would probably be directed towards the red tulip. Furthermore, imagine a single white tulip in a field of daisies. It might be highly salient when viewed perifoveally, that is, in small distance to the current fixation, but might be virtually invisible when viewed in peripheral vision. In comparison to associative accounts, this theory does not imply prior experience so that gaze shifts can simply be driven by the visual input. First empirical evidence for this alternative explanation in the field of action understanding comes from a study showing that adults anticipate the nearest out of several objects being placed in the direction of a movement when they have no further information about the target of the action (Rotman et al., 2006). Notably, regarding the results of Falck-Ytter et al. (2006), it is possible that the salient bucket triggered anticipations in 12-month-olds and adults.

In sum, these alternative explanations provide parsimonious accounts to explain anticipatory looking. This is not to claim that mental states such as goals or intentions cannot influence anticipatory looking, which has clearly been demonstrated in previous studies with adults (Hayhoe et al., 2003; Land, 2009). Though, they seem not to be necessary to anticipate simple movements and might play a minor role in anticipation studies than previously thought. In particular, given the ongoing dispute on the nature of young children's action anticipations, it would be important to directly examine to which extent parsimonious approaches can account for this ability. We experimentally tested this possibility by adapting the setup of Falck-Ytter and colleagues (2006) given their prominent support for motor simulation theories.

To distinguish between simulation and lower level accounts it seems essential to introduce at least two visually identical targets which are both placed within the general direction of the movement and which differ according to their motor kinematics. Notably, different actions are characterized by different movement kinematics (Barrett et al., 2008; Gottwald et al., 2017; Hofsten & Rönnqvist, 1988; Zaal & Thelen, 2005). According to simulation theory, perceiving the kinematics of an action should allow for a simulation of how the action unfolds and thus for a correct anticipation of prediction of its target. A critical test to distinguish between the direct-matching account and lower level mechanisms seems to be to provide observers with kinematic cues that vary as a function of the target location while controlling for other visual differences. If the direct-matching account is correct, we would expect that the kinematic cues can be used to anticipate the target. On the other hand, if anticipations are based on lower level mechanisms we would expect anticipations in the direction of the movement with an anticipation bias towards the closest target (Rotman et al., 2006).

The development of the use of kinematic cues for action understanding

Little is known about the development of the use of kinematic cues for action anticipation and prediction in early and middle childhood. Two studies have claimed that infants use kinematic cues for action anticipations (Ambrosini et al., 2013; Filippi & Woodward, 2016). However, a closer examination indicates that the results are rather inconclusive. In the study by Filippi and Woodward (2016) 13-month-olds watched six identical trials of a grasping movement with either a congruent hand shape or an incongruent hand shape (matching or not matching the orientation of a to be grasped rod). Infants only anticipated the target when the kinematic cues were congruent. Noteworthy, the incongruent kinematic cues did not elicit anticipations to the incorrect target, as would have been expected if infants' anticipations would have been based on the kinematic cues. Moreover, they anticipated the actual object the hand was going to grasp in both conditions, that is, irrespective of the kinematic cues, when they grasped the objects themselves before observing the grasping. Thus, the results seem not to allow for the conclusion that infants use kinematic cues for action anticipation. Notably, as participants watched exactly the same grasping in all six trials, they might have learned which object the hand was going to grasp. This would rather speak for the role of perceptual processes. Similarly, Ambrosini et al. (2013) showed 6-, 8-, 10-months-olds, and adults a hand reaching towards one of two different sized objects with either a whole hand grasp, a precision grasp, or a closed fist. They found that participants anticipated the action faster if the hand was pre-shaped (whole hand or precision grasp) compared to when the actor reached for one of the balls with the closed fist. Yet, not being able to grasp with a closed fist is something else than the processing of kinematic cues. Thus, besides valuable contributions of both studies, neither of them found clear evidence for or against infants' use of kinematic cues for action anticipation.

First empirical evidence that children incorporate kinematic cues to verbally reason about others' actions comes from a study by Bello et al. (2014). They presented 3- to 7-year-olds pictures of hand-object interactions with varying motor kinematics. By the age of three, children were able to discriminate actions such as touching or grasping an object. However, deciding whether an object was grasped to be used vs. grasped to be placed somewhere else based on the handgrip turned out to be challenging—even for the older children. Although performance increased with age, only 6- to 7-year-olds reliably used the grip information to interpret the action. This points to a rather late development of the use of kinematic cues when reasoning about the intention of an observed action.

In sum, it is striking that although the manipulation of motor kinematics seems to allow putting simulation theories at test only very little is known about children's use of kinematic cues for action anticipation and prediction. Our study is the first to close this gap by empirically investigating whether children use kinematic cues to anticipate or predict others' actions while excluding visual inequalities of the target objects (by using two objects of the same size and color) and by reducing the influence of actor-target associations (by presenting the grasping of one out of two objects in a pseudo-random order).

The current study

In the current study 3-, 4-, 10-year-olds, and adults observed a hand reaching for one out of two objects. Both objects were located in the movement direction of the reaching hand, one closer and one further away from the initial position of the hand. The motor kinematics of the hand varied depending on whether the hand reached for the close or far object, respectively. In two experiments we assessed whether participants would use the kinematic cues (1) to anticipate and (2) to predict the target the hand was going to grasp.

We aimed at answering two main questions: First, do children base their anticipations on kinematic cues? Direct matching accounts would predict that children should use kinematic cues early in life to anticipate which out of two objects (both being placed within the direction of the grasping movement) an actor is going to grasp. Yet, if lower level perceptual mechanisms guide children's action anticipations we expect them not to make use of the specific kinematic cues to anticipate the target. Instead, we expect them to either randomly anticipate one of the two targets or, when assessing visual anticipations, to show a preference for the nearer target (see Rotman et al., 2006). The second question was: Do children start to use kinematic cues for implicit action anticipation and verbal action prediction at the same age? If the use of kinematic cues is a rather effortful and slow process which develops later in childhood we expect children to initially use kinematic cues in verbal predictions. On the other hand, if kinematic cues are used in a rather automatic fashion from early on in life we expect children to be able to use them for both, action anticipations as well as action predictions.

We chose 3-year-olds as youngest age group as at this age children are able to grasp objects and adapt their motor trajectory for grasping (Gottwald et al., 2017; Gottwald et al., 2019) Thus, motor simulations of the perceived grasping action should be possible. We included 4- and 10-year-olds as further work suggested that by age 4 children develop increased abilities for reflection (Allen et al., 2021; Allen & Bickhard, 2018) and that reasoning about actions based on kinematics further improves across middle childhood (Bello et al., 2014). Finally, we included an adult group as by this age, kinematic cues are processed and used for action understanding. This allowed us to map developmental differences in children's emerging appreciation of kinematic cues, and to test age differences in using these cues for implicit action anticipation (as assessed by visual anticipations) and explicit, verbal action prediction.

Experiment 1: Action anticipation

Method

Sample size.

Previous research found an effect size of d = 1.31 for 12-month-olds and d = 1.84 for adults between proactive gaze movements during observing a human action compared to a non-human action (Falck-Ytter et al., 2006). Following this research, an a priori power analysis for a mixed ANOVA with an α of 0.05, power = 0.80, an effect size of f = 0.33, four groups and two measurements (close/far) in G*Power 3.1.9.2 (Faul et al., 2009) was calculated and revealed a sample size of 80. We thus aimed at assessing at least 80 participants in each of the two experiments (at least 20 per age group).

Participants.

The final sample of Experiment 1 included 98 participants. After the minimum of 20 participants was reached data collection was continued in case there were further interested participants, leading to small variations in the number of participants per age group. The sample comprised 28 3-year-olds (*M* = 3;04, *SD* = 2.25 months, range = 37-47 months, females = 15), 20 4-year-olds (*M* = 4;04, *SD* = 3.02 months, range = 48-57 months, females = 13), 30 10-year-olds (*M* = 10;04, *SD* = 1.62 months, range = 121-126 months, females = 15) and 20 adults (*M* = 24;05, *SD* = 3;06 years, range = 18-29 years, 11 females). Exact birth dates of two 3-year-olds and three adults were not noted due to experimenter error. Additional 3-year-olds (*n* = 1), 4-year-olds (*n* = 4), 10-year-olds (*n* = 2) and adults (*n* = 1) were tested but not included due to lack of eye-tracking data. Participants came from a large city in Germany. Children were recruited from birth records. Informed consent for participation was given by children’s caregivers and adult participants. Parents received travel cost compensation and children were given a small present. Adult participants were recruited from a student population and got course credit or were paid for their participation. Ethic approval was obtained from the local ethics board.

Stimuli.

Stimulus material consisted of a video divided in a familiarization and testing phase. Stimuli were validated asking ten adults to evaluate how foreseeable it is that the hand reaches for one out of the two objects based on the motor kinematics. Answers were given on a 5-point scale ranging from 1 = “not foreseeable at all”, over 2 = “rather not foreseeable”, 3 = “I don’t know”, 4 = “rather foreseeable”, to 5 = “very clearly foreseeable”. Overall, they evaluated the stimuli as foreseeable (*M* = 4.25, *SD* = 0.79).

Familiarization. At the beginning of the familiarization phase a red screen flashed three times while a sound was played to attract participants' attention. Afterwards, a still image with a dark background and a wooden block (the obstacle) lying on a table was presented (Figure 1, left picture). On the right side of the block a hand laid palm downwards on the table directed to the wooden block. On the left side a small blue cube was placed on a sheet of paper. The area above and left to the wooden block was covered by an occluder, i.e., a blue rectangle. This image was used to familiarize participants with the general scenery. The presentation of this still image was shortly interrupted twice (first by a sound accompanied by a black screen then by a sound accompanied by a red screen) to get participants' attention to the screen. As depicted in Figure 1 the hand then started a grasping movement over the obstacle, disappeared behind the occluder, and reappeared above the blue cube. The hand reached for the blue cube but did not lift it. This video sequence was repeated six times with an attention getter after the third and the sixth trial. With the video we ensured that participants understood that the hand reaches for the cube, that it reaches over the obstacle, and, that it is hidden when behind the occluder but reappears to grasp the cube.

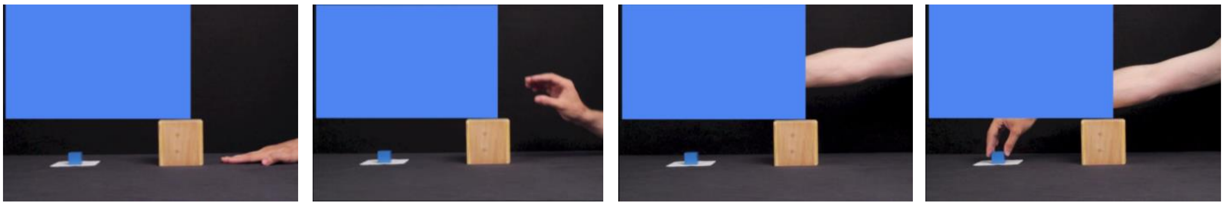


Figure 1. Snapshots of the familiarization video showing the grasping of the single cube.

Test trials. The scene of the testing video differed in one crucial aspect compared to the familiarization video: instead of a single cube on the left side there were two blue cubes presented next to each other: one closer to and one farther away from the hand (Figure 2). The presentation sequence of the still image, black and red screens, and the grasping movement was identical to the familiarization phase. The motor kinematics, however, differed depending on whether the hand reached for the close (Figure 2, top row) or far cube (Figure 2, bottom row), inasmuch as the reaching movement towards the far cube was more pronounced, at greater height, and faster than towards the close cube. For both, the close and the far cube, the hand initiated a reaching movement over the obstacle, disappeared behind the occluder moved a bit further and shortly (1.5s) stopped there. Finally, the hand started to move again behind the occluder and reappeared above one of the cubes and grasped it without lifting it. Between trials a black screen (400ms) was presented. After six trials an attention getter was presented, to maintain participants' attention.

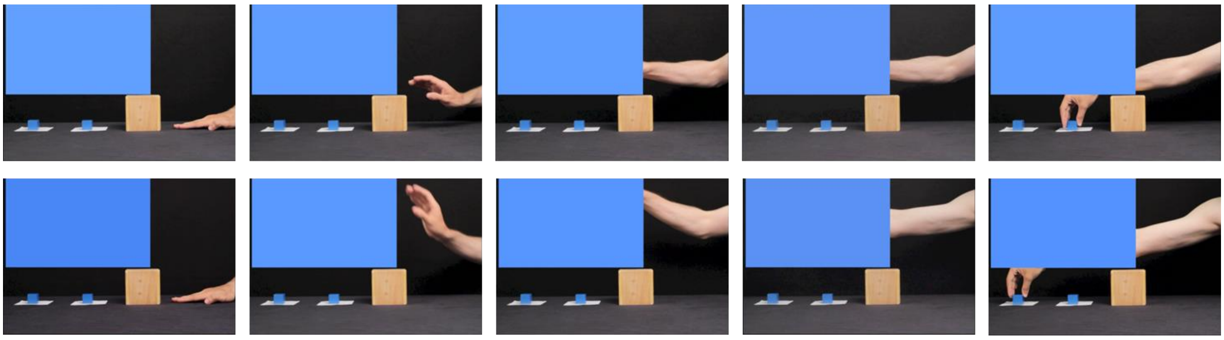


Figure 2. Snapshots of the videos demonstrating the grasping of the close cube (top row) and the far cube (bottom row) during the test phase.

Apparatus and procedure.

Eye movements were recorded with a Tobii T60 eye tracker (60 Hz sampling rate; Tobii Technology, Stockholm, Sweden). Participants sat about 60cm away from an integrated 17″ TFT monitor (1280 × 1024 pixels) on which the stimuli were presented. They were told that they were going to watch a movie. Data was collected and analyzed with Tobii Studio (Tobii Technology, Sweden). After participants had watched the six familiarization trials the testing phase started. Each participant was tested on five close (C) and five far (F) cube trials. Participants were randomly assigned to one of two conditions, in which the close and far cube grasping movements were presented in a pseudo-random order (CFFCCFFCFC or FCCFFCCFCF).

Measures.

The Tobii standard fixation filter with a velocity threshold of 35 pixels/window and a distance threshold of 35 pixels was used to identify fixations. Participants' eye-gazes were measured during the time the hand was behind the occluder. The measurement interval lasted 0.84s (start: hand behind occluder, end: one frame before the hand reappeared) in the familiarization trials and 1.93s in both the close and far testing trials. The measurement interval of the testing trials was chosen to maximize the measurement time for the close trials. Thus, in the close trials the measurement interval started as soon as the hand was behind the occluder and lasted until one frame before the hand reappeared. The movement duration in the far trials was longer than in the close trials. For experimental comparability, the measurement interval in the far trials (in which the hand was behind the occluder for 2.79 s in total) was thus shortened to fit the length of the close trials. As the hand moved faster in the far compared to the close condition, we decided to give participants a bit more time in the beginning to process and anticipate the action. Thus, this led to a measuring interval starting at around 400ms after the hand stopped behind the occluder and ending around 400ms before the hand reappeared behind the occluder. Three areas of interest (AOI) were defined: two goal-AOIs around the target objects, each covering 5.26 % of the screen (Figure 3), and another AOI covering the whole screen (100 %) to control for gazes directed to the screen but to neither of the two goal-AOIs. Similarly, for familiarization trials one AOI covering 5.26% of the screen was defined around the cube and another AOI covered the whole screen. To be included in the analyses, participants had to show eye-gaze data (i.e. fixations to the screen) in at least two of the five test trials for each condition. Trials in which participants showed no eye-gaze data were excluded from further analyses (far: 3-years-old: 13%, 4-years-old: 10%, 10-year-olds: 5%, adults: 4%; close: 3-years-old: 18%, 4-years-old: 9%, 10-year-olds: 5%, adults: 6%).

Three measures were calculated for analyzing participants' gaze behavior, a Frequency Score, a First Fixation Score and a Differential Looking Score (DLS). The former assessed participants' anticipation to the cubes (irrespective of being the correct or incorrect cube) and the latter two assessed participants' expectations about which object the hand is going to grasp.

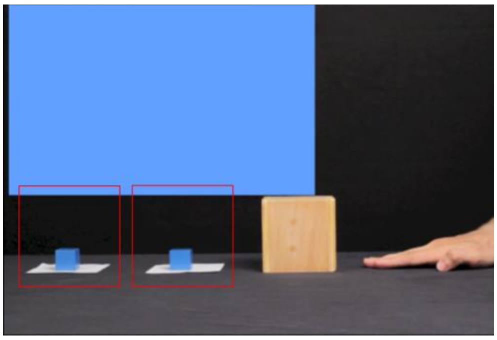


Figure 3. Stimulus material of a test trial. The red boxes indicate the two goal-AOIs.

Frequency of anticipations.

With this score we explored to which extent participants showed anticipations at all—regardless of the correctness of their anticipations. This was done to ensure that our stimuli triggered anticipatory looking to the cubes and to examine whether the age groups differed in their overall number of anticipatory fixations to either of the two cubes. Therefore, anticipations to a cube were coded as 1 and fixations somewhere else on the screen as 0 (similar to Daum et al., 2012; Ganglmayer, Schuwerk, et al., 2019; Hunnius & Bekkering, 2010). Similarly, for the familiarization trials, anticipatory fixations to the single cube were coded as 1 and fixations somewhere else on the screen as 0.

First Fixation Score.

This measure assessed to which of the two goal-AOIs participants fixated first after the hand had disappeared behind the occluder. This measure is well established in the literature (Daum et al., 2012; Ganglmayer, Schuwerk, et al., 2019). For each trial a first fixation to the correct cube, i.e., the cube the hand was going to grasp, was coded as 1 and a fixation to the incorrect cube was coded as 0. If participants showed no fixation to either goal-AOI, that is, they fixated somewhere else on the screen or showed no fixation at all, this was treated as missing value.

Differential Looking Score (DLS).

This score represents the relative looking time to one goal-AOI in relation to the other goal-AOI. We included this measure to account for corrective eye-movements as participants might fixate first on one AOI but direct most of the following fixations to the other AOI (similar to Schuwerk & Paulus, 2016; Senju et al., 2009). The total looking time to the incorrect goal-AOI was subtracted from the total looking time to the correct goal-AOI and divided by the sum of total looking time to both goal-AOIs. This resulted in scores ranging from −1 to 1, with a value towards −1 indicating a looking bias towards the incorrect cube and a value towards 1 a looking bias towards the correct cube.

Analyses strategies.

IBM SPSS Statistics 24 (SPSS Inc., Chicago, IL, USA) was used for statistical analyses. Additionally, Linear Mixed Models (LMMs) and binomial Generalized Linear Mixed Models (GLMMs) were run in R version 3.6.1 using the lmer and glmer functions of the lme4 package.

Frequency of anticipations. To check whether participants anticipate the actions and whether age groups show a comparable number of anticipations, a one-way ANOVA with age group as between factor was calculated for the familiarization and test trials, using the averaged Frequency of Anticipations (averaged over the six trials in the familiarization phase and over the 10 trials in the test phase).

First fixations. To investigate our main question whether anticipations (measured as First Fixations) are based on kinematic cues and to check for differences between age groups, we performed a binomial GLMM with age group, condition (close/far), the interaction age group and condition as fixed effects, participant as random effect, and the First Fixation Score (0/1) as outcome. To examine whether age groups show a comparable number of correct anticipations in each condition, we ran further binomial GLMMs with age group as fixed effect and participant as random effect for each condition separately. To more precisely investigate the differences between age groups, we averaged the First Fixation Score over the five test trials and calculated a one-way ANOVA with age group as between subject factor. Bonferroni corrected pairwise post hoc tests were conducted to determine which age groups differ in their averaged First Fixation Score. Moreover, two-sided *t*-tests were run to test whether participants' First Fixation bias towards one of the goal-AOIs is different from chance.

DLS. In addition to the First Fixations the relative time participants spent looking at one of the two goal-AOIs was analyzed to examine whether participants base their anticipations on the kinematic cues. We ran a mixed ANOVA with condition (close/far) as within-subject factor, age group as between-subject factor, and the DLS averaged over the respective five trials as outcome variable. To investigate whether participants' looking bias would differ significantly between the two categories and from chance level, we ran two-sided *t*-tests for the age groups and categories separately.

Learning effects (DLS). To examine whether participants' looking bias (measured as DLS) over the trials varies between the two categories, we compared a LMM with condition and trial as fixed effects and participant as random effect to a LMM including only trial as fixed effect and participant as random effect. Furthermore, LMMs were run for each age group and condition separately, with participant as random effect, trial as fixed effect and the

DLS of each of the five trials as outcome. To find out whether the looking behavior changed over the trials these models were compared to LMMs that only included the random effect (participant) using Likelihood ratio tests.

Results

Frequency of anticipatory looking.

Familiarization trials.

Participants anticipated in 295 out of 588 trials (50.17%). Table 5 in the Supplementary material shows the frequency of anticipatory looking to the single cube for the six familiarization trials. To analyze whether age groups showed a different number of anticipations to the cube, we calculated an average score over the six trials. A one-way ANOVA with this score and age group as between subject revealed a significant effect of age group, $F(3,93) = 3.76$, $p = .013$, $\eta_p^2 = .11$. Bonferroni corrected pairwise comparisons revealed a significant difference between the adult group ($M = 0.63$, $SE = 0.06$) and 3-year-olds ($M = 0.39$, $SE = 0.06$), $p = .013$. Four-year-olds ($M = 0.54$, $SE = 0.06$) and 10-year-olds ($M = 0.56$, $SE = 0.04$) did not differ significantly from 3-year-olds ($p = .299$ and $p = .102$). Moreover, 4-year-olds and 10-year-olds did not differ significantly from adults or from each other, all *p's* = 1. These analyses show that the stimulus material elicited a comparable number of first fixations to the single cube in all age groups, with only adults showing significantly more anticipations than 3-year-olds.

Test trials.

Participants anticipated in 625 out of 980 trials (63.77%). Table 6 in the Supplementary material shows the descriptive statistics for the frequency of anticipatory looking for each trial, age group, and condition. To analyze whether age groups showed a different number of anticipations to either of the two cubes, we calculated an average score over the ten test trials. A one-way ANOVA with this score and age group as between subject revealed a significant effect of age group, $F(3,94) = 11.45$, $p < .001$, $\eta_p^2 = .27$. Bonferroni corrected pairwise comparisons revealed that adults ($M = 0.83$, $SE = 0.04$) showed a significantly higher number of anticipations compared to 3-year-olds ($M = 0.54$, $SE = 0.04$), $p < .001$, and 4-year-olds ($M = 0.62$, $SE = 0.05$), $p = .003$. Moreover, 10-year-olds ($M = 0.77$, $SE = 0.03$) showed more anticipations than 3-year-olds, $p < .001$. There were no significant differences between 3- and 4-year-olds ($p = 1$), 4- and 10-year-olds ($p = .094$) or 10-year-olds and adults ($p = .798$). These analyses show that the two older age groups anticipated the cubes more often than the younger age groups. More importantly, though, the descriptive statistics indicate that all age groups anticipated the action in all test trials of both categories.

First Fixation Score.

To investigate our main question whether anticipations are based on kinematic cues we analyzed the First Fixation Score. Figure 4 (top) shows the scores of First Fixations averaged over the respective five trials for each age group and condition. A binomial GLMM, with age group, condition (close / far), and the interaction of age group and condition as fixed effects, participant as random effect, and the First Fixation Score (0/1) as outcome, was performed. The interaction of condition and age group turned out to significantly predict the First Fixation Score ($b = 0.88$, $SE = 0.29$, $z = 3.02$, $p = .002$, odds ratio = 2.41, 95% CI[1.38, 4.38]). Moreover, there was a significant effect of condition ($b = -6.27$, $SE = 0.90$, $z = -6.99$, $p < .001$, odds ratio = 0.001, 95% CI[0.0002, 0.009]) whereas age group was not a significant predictor ($b = -0.10$, $SE = 0.26$, $z = -0.38$, $p = .701$, odds ratio = 0.90, 95% CI[0.53, 1.49]).

To further investigate the interaction effect, binomial GLMMs with age group as fixed effects and participant as random effect were run for each condition separately.

For the close condition age group ($b = -0.10$, $SE = .31$, $z = -0.33$, $p = .743$, odds ratio = 0.90, 95% CI[0.35, 0.85]) did not significantly predict the First Fixation Score. This indicates that over all trials all age groups showed a comparable number of correct anticipations in the close condition. One-sample *t*-tests against chance revealed that all age groups anticipated to the close cube above chance (Table 1).

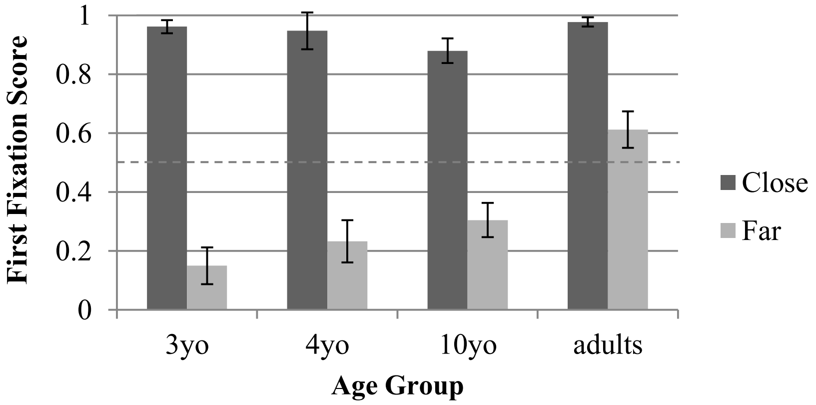
For the far condition, on the other hand, age group ($b = 0.85$, $SE = .19$, $z = 4.39$, $p < .001$, odds ratio = 2.34, 95% CI[1.64, 3.57]) significantly predicted the First Fixation Score. We thus averaged the First Fixation Score over the five far trials and calculated a one-way ANOVA with age group as between subject factor. The ANOVA revealed a significant result for age group $F(3,89) = 9.13$, $p < .001$ $\eta_p^2 = .23$. Bonferroni corrected post hoc tests revealed significant differences between adults ($M = .61$, $SE = .06$) and all other age groups: 3-year-olds ($M = .15$, $SE = .06$), $p < .001$, 4-year-olds ($M = .23$, $SE = .07$), $p = .001$, and 10-year-olds ($M = .30$, $SE = .06$), $p = .005$. Children age groups did not perform significantly different from each other (all *p's* ≥ .400). This shows that in the far condition adults anticipated to the correct far cube more often than the children groups. Except for adults, all one-sample *t*-tests against chance level for all age groups were significant (Table 1), indicating that children's performance was significantly below chance level in the far trials. Hence, children rather showed first fixations to the incorrect close cube in the far condition, whereas only adults showed a tendency of looking towards the correct far cube, even though being not significant.

DLS.

To account for corrective eye movements, for which the First Fixation Score does not account for, the DLS was analyzed. Figure 4 (bottom) shows the mean DLS for each age group and condition. The mixed ANOVA resulted in a significant interaction effect between condition (close/far) and age group, $F(3, 94) = 5.79$, $p = .001$, $\eta_p^2 = .16$. To further investigate this effect, paired-sample *t*-tests for each age group between the far and close condition were performed. All age groups performed significantly better in the close compared to the far condition (3-year-olds: $t(27) = 10.17$, $p < .001$; 4-year-olds: $t(19) = 7.49$, $p < .001$; 10-year-olds: $t(29) = 8.41$, $p < .001$; adults: $t(19) = 2.43$, $p = .025$).

A one-way ANOVA for the DLS of the close condition and age group as between-subject factor revealed no effect of age group, *F*(3, 94) = 2.06, *p* = .110, $\eta_p^2 = .06$, indicating that all four age groups performed similarly in the close trials. One-sample *t*-tests against chance revealed that—similar to the First Fixation Score—all age groups anticipated the correct close cube above chance (Table 1).

An ANOVA for the DLS of the far condition resulted in a significant effect of age group, *F*(3, 94) = 18.91, *p* < .001, $\eta_p^2 = .38$. Bonferroni corrected pairwise comparisons revealed a significant difference between adults (*M* = 0.39, *SE* = 0.10) and 3-year-olds (*M* = -0.32, *SE* = 0.05), as well as between adults and 4-year-olds (*M* = -0.31, *SE* = 0.08), and adults and 10-year-olds (M = -0.24, SE = 0.07), all *p*'s < .001. As for the First Fixation Score, the children groups did not differ significantly from each other (all *p's* = 1). Thus, adults looked longer at the correct cube than the other age groups when the hand reached for the far object. One-sample *t*-tests against chance for each age group revealed a similar pattern of result as for the First Fixation Score: Only adults showed a looking bias towards the correct far cube, whereas all children groups looked longer on the incorrect close cube.



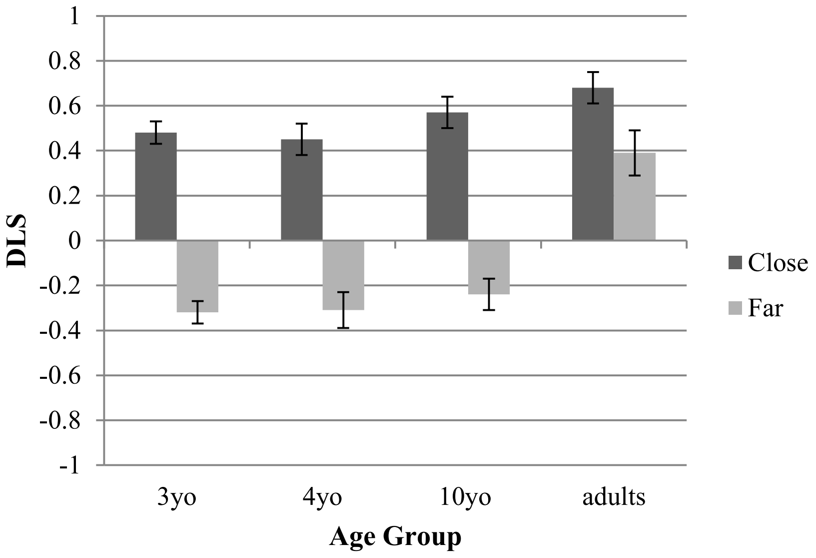


Figure 4. First Fixation Score (top) and Differential Looking Score (DLS) (bottom) for each age group and condition. Dashed line represents chance level at 0.5 for the First Fixation Score. Error bars depict standard errors.

Table 1. *Results of the one-sample t-tests against chance level for the First Fixation Score (chance level at 0.5) and DLS (chance level at 0) for each age group and condition (close/far).*

		3-year-olds		**4-year-olds**		**10-year-olds**		**adults**	
		T	*p*	*T*	*p*	*T*	*p*	*T*	*p*
First Fixation Score									
	Close	20.53	< .001	8.50	< .001	9.10	< .001	30.63	< .001
	Far	-5.60	< .001	-3.73	.002	-3.37	.002	1.81	.086
DLS									
	Close	8.90	< .001	6.10	< .001	8.41	< .001	10.19	< .001
	Far	-6.12	< .001	-4.00	.001	-3.50	.002	3.86	.001

Note: DLS = Differential Looking Score

Analyses of learning effects (DLS).

The DLS over the five trials for each age group and condition are depicted in Figure 6 in the Supplementary material. An LMM with condition and trial as fixed effects and participant as random effect (full model) was compared to a model including only trial as fixed effect and participant as random effect (reduced model). A Likelihood ratio test revealed that the model that additionally included condition as predictor matched the data significantly better (AIC reduced model = 1920.49, AIC full model = 1653.04, $\chi^2(1) = 269.45$, $p < .001$). Thus, all further analyses were computed for the close and far condition separately.

To test for learning effects over the five trials Linear Mixed Models (LMM) with participant as random effect, trial as fixed effect and DLS as outcome were calculated (Table 2). Moreover, we compared a model including the random and fixed effect to a model including only the random effect for each age group and condition (Likelihood ratio test in Table 2). For the 10-year-olds in the far condition the model additionally including trial as a predictor turned out to fit the data significantly better than the model including only participant as a random effect. None of the other model comparisons revealed a significant difference. This indicates that none of the age groups showed a change in looking bias over the five trials in the close condition and that only 10-year-olds showed a learning effect in the far trials. Nevertheless, 10-year-olds DLS in the fifth trial ($M = .13$, $SE = .16$) is at chance level, $t(26) = 0.87$, $p = .393$. This indicates that even though 10-year-olds' looking bias towards the correct far cube increased over the five trials, they still performed at chance in the last trial.

Table 2. *Results of the Linear Mixed Models over the five trials for the DLS and of the Likelihood ratio tests.*

Age	**Cond.**	**Intercept**	**Trial**	**Likelihood ratio test**

group		*β*	*SE*	*T*	*β*	*SE*	*T*	AIC$_{\text{rand}}$	AIC$_{\text{full}}$	χ^2	*p*
3-year-olds	Close	0.49	0.11	4.45	0.00	0.03	-0.10	185.33	187.32	0.01	.920
	Far	-0.49	0.10	-4.71	0.06	0.03	1.78	183.11	181.94	3.17	.075
4-year-olds	Close	0.46	0.13	3.57	0.00	0.04	-0.06	140.67	142.67	0.00	.949
	Far	-0.39	0.16	-2.47	0.03	0.05	0.57	177.46	179.13	0.32	.570
10-year-olds	Close	0.47	0.11	4.28	0.03	0.03	1.08	241.46	242.28	1.18	.277
	Far	-0.51	0.14	-3.62	0.09	0.04	2.22	319.57	316.70	4.88	.027
Adults	Close	0.56	0.11	5.12	0.04	0.03	1.42	121.71	121.70	2.01	.157
	Far	0.16	0.16	0.98	0.07	0.04	1.67	203.63	202.87	2.76	.097

Note: The likelihood ratio test compares a model including only the random effect (participant) with the full model including the fixed (trial) and random effect (participant). AIC = Akaike Information Criterion.

Discussion

The first experiment aimed at investigating whether kinematic cues allow to anticipate the target of an observed action, as proposed by the direct-matching hypothesis and other simulation theories (e.g., Gallese et al., 2004; Rizzolatti et al., 2001; Rizzolatti & Sinigaglia, 2010), or whether action anticipations are rather based on lower level perceptual mechanisms (Ganglmayer, Attig, et al., 2019; Hunnius & Bekkering, 2014; Ruffman et al., 2012). To this end, participants watched a video in which the motor kinematics of a reaching movement varied depending on whether the actor grasped a close or far object. We assessed whether participants would use the motor kinematics to anticipate which of the two target-objects the actor was going to grasp. Results of both key measures, the First Fixation Score and the Differential Looking Score, provide evidence that only adults but not 3- to 10-year-olds base their anticipations on kinematic cues, speaking against simulation theories (at least for early and middle childhood). Instead, the findings support claims that action anticipations are rather based on early developing perceptual mechanisms.

The key finding was that all child groups showed a strong looking bias towards the close cube in the far condition, indicating that they did not base their visual anticipations on the kinematic cues. Whereas the direct-matching and lower level perceptual mechanism accounts make similar predictions for the close condition (i.e., anticipations to the close cube), the far condition constitutes a crucial test to differentiate between the two approaches. The direct-matching account predicts anticipations to the far object (based on the kinematic cues) whereas lower level perceptual mechanisms predict anticipations to the close cube (based on anticipating the direction of the movement until coming across an object). Crucially, instead of using the movement kinematics to anticipate the target, children showed a looking bias towards the close cube in both conditions. Thus, the results speak against simulation theories, according to which observing an action allows to simulate and consequently to anticipate the action goal (e.g., Gallese et al., 2004; Goldman, 2006; Rizzolatti et al., 2001). Instead, children's looking bias towards the close cube—irrespective of the kinematic cues—supports claims that action anticipations are based on lower level perceptual mechanisms. Observing the grasping movement elicited anticipations to the close object. Though, it did not allow for differentiating which of the objects the hand was going to grasp. Instead, children anticipated the closer cube, which was presumably more salient than the far object due to the lower visual eccentricity. This gaze pattern was present in all children age groups.

Interestingly, 10-year-olds improved in anticipating the grasping of the far cube over the trials. Thus, observing the action repeatedly contributed to improvements in action anticipating. This points to the importance of statistical learning for visually anticipating other's actions (e.g., Ruffman et al., 2012; Ruffman, 2014). A more detailed discussion of this finding can be found in the General Discussion.

Two measures were used to assess action anticipations and, thus, to reveal participants' expectations about the actor's action: the First Fixation Score and the Differential Looking Score. The First Fixation Score constitutes a more stringent measure of action anticipation as it only includes the first anticipatory fixation to one of the two target-objects. The DLS on the other hand considers all fixations during the anticipatory period and is thus also sensitive to corrective eye movements. Noteworthy, both measures showed a similar pattern of results. Thus, the results cannot be explained by an initially fast but inaccurate eye movement which is subsequently corrected. Instead, both measures indicate that children did not use kinematic cues for target anticipation.

Some alternative explanations are considered rather unlikely. First, participants were provided with no other information about the actor's goal than the kinematic cues. Particularly, the single cube in the familiarization phase was located centrally between the two cubes of the testing phase. Thus, it is unlikely that the familiarization phase enhanced anticipations to either of the two objects. Second, the analyses of the frequency of anticipatory looking during the familiarization and test phase proved that the grasping movement elicited anticipatory fixations in all age groups. Although, there were differences in the number of anticipatory fixations between age groups, even 3-year-olds anticipated the grasping actions. Thus, it is unlikely that children's looking bias towards the incorrect close cube in the far test trials can merely be explained by less overall anticipatory fixations in children compared to adults.

Experiment 1 showed that 3- to 10-year-olds anticipate towards an object in the general direction of a grasping movement but do not base their anticipations on kinematic cues. This indicates that their anticipations are rather based on lower level perceptual mechanisms than on matching the observed movement. Moreover, it suggests that kinematic cues are not processed automatically from early on in life. Stronger evidence that the processing of kinematic cues is not an automatic process but relies (at least initially) on slow and cognitively demanding processes would be provided by demonstrating that children use kinematic cues for action prediction before they use them for action anticipations. To examine this possibility we conducted Experiment 2.

Experiment 2: Action Prediction

Method

Participants.

The final sample comprised 80 participants, 20 for each of four age groups: 3-year-olds (*M* = 3;07, 11 females), 4-year-olds (*M* = 4;07, 9 females), 10-year-olds (*M* = 10;04, 11 females), and adults (*M* = 25;03, 11 females). Per age group 20 analyzable participants served as a stop criterion for data collection. Due to an experimenter error, this criterion was neglected for the 10-year-olds. Thus, we excluded 15 additionally tested 10-year-olds to ensure comparable sample sizes across age groups[1]. Additional 11 participants were excluded due to providing too few trials (3-year-olds: *n* = 6), refusing to participate in the task (3-year-olds: *n* = 1), or technical issues (3-year-olds: *n* = 3, 4-year-olds: *n* = 1). Participants came from the same population as in Experiment 1, received the same compensation for participation, and gave informed consent about participation as described in Experiment 1. Ethic approval was obtained from the local ethics board.

Stimuli.

Stimulus material was identical to Experiment 1.

Setting and procedure.

Participants sat in front of a screen and were told that they were going to watch a movie. The procedure was identical to that of Experiment 1 and differed only in two crucial aspects: 1) there was no eye tracking and 2) the experimenter stopped the video when the hand was behind the occluder and asked the participant “Show me, which cube grasps the

---

[1] All analyses were additionally run including the 15 10-year-olds. There were no differences in the results, except for the learning effects in the close condition over the 1st to 5th test trials as well as over the 2nd to 5th test trial being now significant for the 10-year-olds.

hand?”. Otherwise, the procedure remained the same so that after indicating which cube the hand was going to grasp participants saw which cube the hand actually grasped.

Explicit prediction score.

Each trial was coded as 1 if participants indicated the cube the actor was going to grasp and as 0 if they indicated the wrong cube. Trials in which participants pointed to any other location than to the target cubes or indicated that they do not know the answer were considered as invalid and excluded from the analyses (close: 3-years-old: 7%, 4-years-old: 1%, 10-years-old: 1%; far: 3-years-old: 2%, 4-years-old: 3%). Since this was only a small proportion for all age groups we did not further analyze differences in the general number of predictions.

Analyses strategy.

We used the same programs and functions as in the first experiment. To analyze whether participants’ verbal predictions differ in the close compared to the far trials and to check for differences between age groups, a binomial GLMM with age group, condition (close / far), and the interaction of age group and condition as fixed effects, participant as random effect, and the Prediction Score (0/1) as outcome was performed. To further examine differences between age groups we averaged the Prediction Score over all trials and performed a one-way ANOVA with age group as between factor and subsequent Bonferroni post hoc tests. By using one-sample *t*-tests against chance we tested whether the age groups predicted the correct and incorrect cube differently from chance.

To statistically examine the learning effects over the five trials for each age group and condition binomial GLMMs with participant as random effect, trial as fixed effect, and the Prediction Score (0/1) as outcome were calculated. As participants did not know that they were going to be asked to predict the target of the action in the first trial, we rerun the learning effect analyses for the 2nd to 5th trial.

Results

Examining the data revealed one extreme outlier (below 1st quartile – 3*interquartile range) in the adult sample. This participant was excluded from further analyses.

Explicit prediction score.

Figure 5 shows the averaged Explicit Prediction Score over the respective five trials of each condition (close/far) and for each age group. A binomial GLMM with age group, condition (close/far), and the interaction age group and condition as fixed effects, participant as random effect, and the Prediction Score (0/1) as outcome was performed. The interaction of condition and age group did not significantly predict participants' Prediction Score (b = -0.09, SE = 0.20, z = -0.44, p = .661, odds ratio = 0.91, 95% CI[0.61, 1.36]). There was a main effect of age group (b = 1.21, SE = 0.17, z = 7.22, p < .001, odds ratio = 3.36, 95% CI[2.46, 4.79]), but no main effect of condition (b = -0.01, SE = 0.43, z = -0.04, p = .971, odds ratio = 0.98, 95% CI[0.42, 2.31]). We thus averaged the Prediction Score over all 10 test trials and ran a one-way ANOVA with age group as between subject factor to further examine the age group differences. The ANOVA confirmed the main effect of age group $F(3,75) = 85.41$, p < .001, η_p^2 = .77. Post hoc Bonferroni corrected comparisons revealed that 3-year-olds (M = .52, SE = .03) and 4-year-olds (M = .50, SE = .04) performed significantly worse compared to 10-year-olds (M = .92, SE = .02) and adults (M = .96, SE = .02), all *p's* < .001. There was no statistically significant difference between 3-year-olds and 4-year-olds or 10-year-olds and adults, respectively (both *p's* = 1.0). This indicates that over both categories the two younger children groups predicted the cube the hand was going to grasp significantly worse than 10-year-olds and adults.

One-sample t-tests against chance (here: 0.5) for each age group revealed that neither 3-year-olds ($t(19)$ = .51, p =.613) nor 4-year-olds ($t(19)$ = .13, p =.896) performed differently from chance level. Ten-year-olds (*t*(19) = 24.37, *p* < .001) and adults (*t*(18) = 26.53, *p* < .001) on the other hand performed above chance.

In sum, the results show that only 10-year-olds and adults correctly predicted the cube the hand was going to grasp. There was no difference in predicting the close compared to the far trials.

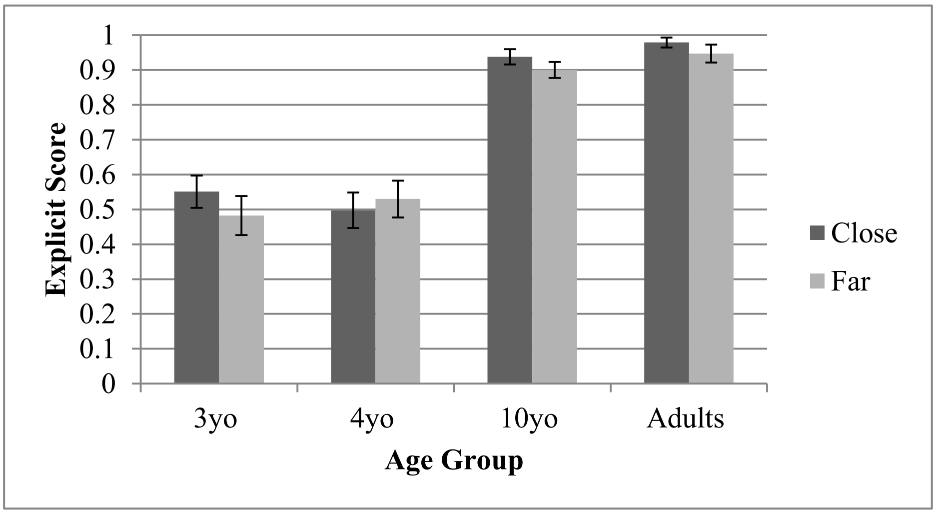


Figure 5. Explicit Close and Far Scores for the four age groups. Chance level at 0.5. Error bars depict standard errors.

Analyses of learning effects.

The Prediction Scores over the five trials for each condition and age group are depicted in Figure 7 in the Supplementary material. To statistically examine learning effects over the five trials of each condition binomial GLMMs with participant as random effect, trial as fixed effect and the Prediction Score (0/1) as outcome were calculated for each age group (Table 3). Trial turned out to be a significant predictor for the 10-year-olds in the far condition. For the close condition, trial was marginally significant for the 10-year-old. None of the other age groups showed a change in correctly predicting the grasping movement over the five trials in either the close or the far condition.

Table 3. *Results of the Generalized Linear Mixed Models over the five trials for each condition and age group.*

Age group	**Cond.**	**Intercept**				**Slope**				**R^2**
		β	*SE*	*z*	*p*	*β*	*SE*	*z*	*p*	
3-year-olds	Close	-0.36	0.49	-0.74	.461	0.19	0.15	1.24	.214	.02
	Far	0.34	0.50	0.68	.497	-0.14	0.15	-0.95	.344	.07
4-year-olds	Close	-0.52	0.48	-1.08	.282	0.16	0.14	1.14	.254	.02
	Far	-0.22	0.48	-0.45	.653	0.09	0.14	0.64	.524	.00
10-year-olds	Close	0.88	0.87	1.01	.311	0.77	0.40	1.93	.054	.26
	Far	0.01	0.72	0.02	.987	0.93	0.35	2.68	.007	.35
Adults	Close	2.40	1.43	1.68	.092	0.58	0.61	0.95	.343	.17
	Far	3.40	2.51	1.35	.177	3.05	1.82	1.67	.094	.96

Post hoc analyses

Noteworthy, participants did not know that they were going to be asked to predict the target of the action in the first trial. As this might be crucial to trigger the explicit information processing system we decided to further analyze the 2nd to 5th trial.

To statistically test for performance changes over the 2nd to 5th trial further binomial GLMMs with the same variables as before were calculated (Table 4). Trial turned out to be a significant predictor for 4-year-olds' performance in the far condition, indicating a learning effect over the four trials. For the close condition trial was marginally significant for the 4-year-olds. Except for the 4-year-olds, none of the other age groups showed a learning effect in either the close or the far condition over the second to fifth trial.

Table 4. *Results of the Generalized Linear Mixed Models over the 2nd to 5th trial for each condition and all age groups.*

Age group	**Cond.**	**Intercept**				**Slope**				**R^2**
		β	*SE*	*z*	*p*	*β*	*SE*	*z*	*p*	
3-year-	Close	-0.99	0.80	-1.23	.217	0.34	0.22	1.56	.117	.04

olds	Far	-0.81	0.79	-1.02	.307	0.15	0.21	0.69	.489	.06
4-year-olds	Close	-1.15	0.76	-1.85	.064	0.39	0.21	1.88	.061	.05
	Far	-2.02	0.81	-2.49	.013	0.54	0.22	2.48	.013	.10
10-year-olds	Close	0.77	1.88	0.41	.681	0.81	0.66	1.22	.224	.20
	Far	0.72	1.62	0.45	.654	0.71	0.55	1.29	.195	.16
Adults	Close	2.96	3.08	0.95	.340	0.42	0.97	0.44	.660	.06
	Far	2.94	3.08	0.95	.340	0.42	0.97	0.44	.660	.06

Discussion

To investigate whether 3- to 10-year-old children and adults use kinematic cues to explicitly predict others’ actions we repeated Experiment 1 with one crucial change: this time we asked participants to predict which cube the hand was going to grasp. Importantly, even in this context, 3- and 4-year-old children did not use kinematic cues to predict the action. Though, 10-year-olds used the kinematic cues to correctly predict the target-object, whereas they did not correctly anticipate the target-object in the first experiment. This suggests that children use kinematic cues for explicit action prediction before using them for action anticipations. Thus, using kinematic cues to predict others’ actions seems to require—at least initially—some explicit reasoning, indicating that it is a rather effortful and slow process.

Interestingly, 4-year-olds and 10-year-olds showed a learning effect over the trials. Whereas 10-year-olds’ learning effect was present over the first to fifth trial it was not significant anymore over the second to fifth trial, suggesting that it can be traced back to the increase in performance between the first and second trial. This indicates that the explicit reasoning after the first test trial was crucial for 10-year-olds’ increase in prediction performance. The results showed a slightly different pattern for 4-year-olds. Four-year-olds seemed to rely on the repeated visual feedback about which cube the actor grasped and adapted their predictions accordingly over time. Thus, 4-year-olds learned to use the kinematic cues to predict the target over the trials. This suggests that children by the age of four can learn to use kinematic cues to predict others’ actions.

It might be argued that language deficits along with deficits in understanding the task in the younger age groups account for children's poor performance. We consider this unlikely. First, numerous other studies have shown that children are able to verbally predict actions by the age of three (e.g., Paulus et al., 2017; Wellman & Woolley, 1990). Second, in the current experiment all age groups predicted the grasping movement in the vast majority of the trials (even though incorrectly). Only few invalid trials had to be excluded from the analyses. This indicates that already 3-year-olds were engaged in predicting the grasping action but that they were just not able to take the movement kinematics into account.

General Discussion

One key question concerns how children come to understand and predict others behavior (e.g., Carpendale & Lewis, 2004; Hunnius & Bekkering, 2014; Ruffman et al., 2012). Simulation theories have claimed that we understand others and anticipate their behaviors by activating mental processes that would evoke similar behaviors in ourselves. Consequently, it has been argued that simulating others constitutes the basis of early social cognition. One approach, the direct-matching hypothesis of action understanding, claims that we understand others actions by matching an observed action onto our own motor repertoire (Gallese et al., 2004; Rizzolatti et al., 2001; Rizzolatti & Sinigaglia, 2010) and that children are thus able to anticipate an action if they master the action themselves (Falck-Ytter et al., 2006; Kanakogi & Itakura, 2011). Though, some researchers have challenged this view (e.g., Southgate, 2013). Others have suggested that action anticipations are rather based on lower level perceptual mechanisms, leading to an ongoing debate about the underlying mechanisms of action anticipation (Ganglmayer, Attig, et al., 2019; Hunnius & Bekkering, 2014; Paulus, 2012; Ruffman et al., 2012). The current study contributes to this debate by showing that 3- to 10-year-olds do not use kinematic cues to anticipate the target of an action but instead anticipate the nearest object placed in the general movement direction of the grasping hand.

Thus, the results speak against direct-matching accounts, according to which kinematic cues allow to anticipate the action target. Instead, the findings provide evidence for claims that action anticipations are based on lower level perceptual mechanisms. This will be discussed in more detail in the following paragraphs.

As shown in Experiment 1, only adults but not 3-to10-year-olds based their visual anticipations on kinematic cues. According to the direct-matching hypothesis others' actions are predicted by matching an observed action onto one's own motor repertoire (Gallese et al., 2004; Rizzolatti et al., 2001; Rizzolatti & Sinigaglia, 2010). In turn, having a motor representation of a certain action allows for anticipating the action when observing someone else performing it. Noteworthy, it is undoubtable that by three years, children can reach and grasp objects. Even infants and toddlers show different movement kinematics during reaching and grasping actions, depending on the location and characteristics of the target objects as well as on the subsequently to be performed action (Chen et al., 2010; Claxton et al., 2003; Gottwald et al., 2017; Zaal & Thelen, 2005). For example, Gottwald et al. (2017) showed that the longer the distance between a to-be-grasped object and its target location, the slower infants' initial reaching towards the to-be-grasped object. Given this set of findings, it is very unlikely that children by the age of three did not have the respective motor representations of the observed actions. Nevertheless, 3- to 10-year-olds were not able to use the kinematic cues to anticipate the target-object, speaking against simulation accounts.

There is a recent debate whether visual anticipations might rather be based on lower level perceptual mechanisms such as statistical and associative learning (Daum et al., 2012; Ganglmayer, Attig, et al., 2019; Hunnius & Bekkering, 2014; Ruffman et al., 2012). Moreover, it has been claimed that gaze shifts can be triggered by non-foveal retinal stimulation (Harris, 1989). The current study supports these views by showing that children until the age of 10 years anticipate the first object they come across when following the general direction of the grasping movement—irrespective of the specific kinematic cues. By the age of three children have observed many grasping movements and by extracting regularities across several observations they might have learned that a grasping movement usually continues in its initial direction and is directed at an object. Thus, observing a grasping movement evokes anticipations to the first (i.e., the nearest) object they come across when following the movement direction. Related evidence for this explanation comes from computational models suggesting that the recognition of biological movements is based on learned prototypical visual patterns (Giese & Poggio, 2003).

To which extend observing experience is crucial to elicit anticipations remains to be seen in future studies. Indeed, it might be that the anticipations found in the current study can be explained by even more simple mechanisms of non-foveal retinal stimulation (Harris, 1989). Importantly, there is first empirical evidence supporting this view as an equivalent default strategy has been found in adults. When anticipating unpredictable grasping actions adults anticipated towards a closer compared to a further away object (Rotman et al., 2006).

In any case, anticipating the nearest object in the movement direction is an efficient strategy, as it is “on the way” to the following objects anyway. Thus, shifting gaze to the next object only if the currently fixated object is not located in the movement direction anymore (e.g., the grasping hand has passed the nearest object) allows for short distanced gaze shifts while avoiding missing potential targets.

Interestingly, 10-year-olds showed a learning effect of anticipating the grasping of the far object over the trials. Thus, repeated observations might eventually lead to associations between movement kinematics of specific reaching movements (e.g., reaching for a close vs. far object) and the location at which the grasping is directed at. This points to an important role of statistical learning in action anticipation (Monroy et al., 2018, 2017; Monroy, Meyer, et al., 2017) and relates to claims of its role as a domain general learning mechanism

(Kirkham et al., 2002; Kirkham et al., 2007). Taken together, our findings support theoretical views suggesting that lower level perceptual mechanisms seem to play an important role for visual action anticipations, particularly in early development, whereas kinematic cues seem less relevant.

Importantly, such lower level mechanisms might also account for the results of previous anticipation studies and, indeed, previous studies have shown that motor activation during action observation is not a prerequisite for action understanding but rather follows action interpretation, speaking against simulation accounts (Pomiechowska & Csibra, 2017). However, lower level explanations have often been neglected in favor of interpretations based on simulation accounts (e.g., Falck-Ytter et al., 2006; Kanakogi & Itakura, 2011). For example, Falck-Ytter et al. (2006) showed that 12-month-olds and adults but not 6-month-olds visually anticipate a bucket in which an actor is about to place a toy. They concluded that their findings support the direct-matching hypothesis as only 12-month-olds and adults but not 6-month-olds master such an action themselves and were therefore able to anticipate the target of the action. However, participants were presented with a single, highly salient target-object (a red bucket with a 3D happy face connected to it) in the direction of the placing movement of the hand. Thus, the alternative explanation that participants followed the general direction of the placing movement and anticipated the only available object that was placed within this region cannot be ruled out. The current study overcomes this limitation and shows that action anticipations are not based on mirroring motor kinematics but are rather explained by lower level perceptual mechanisms.

Besides, many studies have used action anticipation and other non-verbal measures to examine the early development of understanding others actions (Cannon & Woodward, 2012; Daum et al., 2012; Falck-Ytter et al., 2006; Gredebäck et al., 2009; Monroy et al., 2020). This line of research has led to the predominant view that children show an implicit understanding of others' actions before being able to verbally reason about others' actions. However, recent studies indicate that this is not necessarily the case. Predicting other's actions sometimes precedes correctly anticipating the same action (e.g., Paulus et al., 2017; Schuwerk & Paulus, 2016). These findings have fueled an ongoing debate about whether different processes might underlie the accomplishment of implicit and explicit action understanding (Barone et al., 2019; Grosse Wiesmann et al., 2017; Paulus et al., 2017). Our study contributes to this debate by demonstrating that kinematic cues are initially used for action prediction rather than for action anticipation. This supports claims that children do not necessarily show an implicit understanding of others' actions before showing an explicit understanding.

Noteworthy, 10-year-olds used kinematic cues to predict the action target and 4-year-olds learned to do so over the trials, whereas neither of the age groups used kinematic cues to correctly anticipate the action in the first experiment. Thus, kinematic cues are earlier used for explicit action predictions before being used for action anticipations. This is interesting in two ways: First, it shows that children's inability to use kinematic cues in the first experiment is not due to the nature of the used kinematic cues. In other words, the kinematic cues provided sufficient information to enable participants to form an expectation about which of the two cubes the hand was going to grasp. Second, and even more important, the findings suggest that kinematic cues are not processed automatically from early on in life, as would be expected by direct-matching accounts (e.g., Gallese et al., 2004; Rizzolatti et al., 2001). Instead, the development of the use of kinematic cues seems to initially rely on rather cognitively demanding and slow processes. This adds to considerations that two systems might underlie social information processing (e.g., Apperly & Butterfill, 2009; Strack & Deutsch, 2004) and challenges the predominant view that children generally show an implicit understanding of others' actions before showing an explicit understanding. Whereas previous research indicated that sometimes children are able to predict another's action before they correctly anticipate the same action, the current study is the first to show that this holds true for the use of kinematic cues.

Despite the interesting findings, the current study also has some limitations. We focused on four age groups with (partially) considerable age differences. This was based on theoretical considerations that an assessment of these age groups would be particularly informative. Our results revealed a significant improvement between 4 and 10 years of age when verbally predicting other's grasping actions and after 10 years of age when visually anticipating such actions. This calls for future empirical work that focuses specifically on these age ranges.

Moreover, future studies should extend our results using various kinds of kinematic cue (for example, using different types of grips) in order to explore whether different kinematic and motor cues might be processed earlier in development. When designing new studies special attention should be paid to control for perceptual differences that might influence anticipations. One confounding factor could be differences in salience of the target objects. Indeed, previous studies indicate that adults as well as infants anticipate towards more salient bigger objects compared to smaller ones (Ambrosini et al., 2011; Ambrosini et al., 2013). Thus, stimuli material needs to be designed carefully to ensure that anticipations are based on the kinematic cues rather than on other visual features.

To avoid an interaction of action anticipations and action predictions we ran the two experiments with separate samples. It could be interesting to examine the interplay of action anticipations and predictions in more detail in future studies. Noteworthy, a recent study found that the implicit and explicit system seem to inform each other by 3 years of age (Paulus et al., 2017). Whether an explicit instruction like "which hand is the actor going to grasp?" might strengthen learning effects in visual anticipations remains an open question for further research.

Taken together, the current study adds to two ongoing debates: First, whether different processes underlie visual action anticipations and verbal action predictions and, second, whether action anticipations are based on lower level perceptual mechanisms rather than on simulating others' actions. We showed that children use kinematic cues first to predict other's actions before using them to anticipate other's actions. This suggests that the use of kinematic cues requires—at least initially—some explicit reasoning, indicating that it is a rather cognitively demanding and slow process. Furthermore, children did not base their anticipations on the specific movement kinematics of the observed actions. This provides evidence that action anticipations are guided by lower level perceptual mechanisms rather than by simulating others' actions.

References

Allen, J. W., & Bickhard, M. H. (2018). Stage fright: Internal reflection as a domain general enabling constraint on the emergence of explicit thought. *Cognitive Development*, *45*, 77–91. https://doi.org/10.1016/j.cogdev.2017.12.005

Allen, J. W., Çelik, B., & Bickhard, M. H. (2021). Age 4 transitions: Reflection as a domain-general development for explicit reasoning. *Cognitive Development*, *59*, 101091. https://doi.org/10.1016/j.cogdev.2021.101091

Ambrosini, E., Costantini, M., & Sinigaglia, C. (2011). Grasping with the eyes. *Journal of Neurophysiology*, *106*(3), 1437–1442. https://doi.org/10.1152/jn.00118.2011

Ambrosini, E., Reddy, V., Looper, A. de, Costantini, M., Lopez, B., & Sinigaglia, C. (2013). Looking ahead: Anticipatory gaze and motor ability in infancy. *PloS ONE*, *8*(7), e67916. https://doi.org/10.1371/journal.pone.0067916

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, *116*(4), 953–970. https://doi.org/10.1037/a0016923

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, *14*(3), 110–118. https://doi.org/10.1016/j.tics.2009.12.006

Barone, P., Corradi, G., & Gomila, A. (2019). Infants' performance in spontaneous-response false belief tasks: A review and meta-analysis. *Infant Behavior & Development*, *57*, 101350. https://doi.org/10.1016/j.infbeh.2019.101350

Barrett, T. M., Traupman, E., & Needham, A. (2008). Infants' visual anticipation of object structure in grasp planning. *Infant Behavior & Development*, *31*(1), 1–9. https://doi.org/10.1016/j.infbeh.2007.05.004

Bartsch, K., & Wellman, H. M. (1989). Young children's attribution of action to beliefs and desires. *Child Development*, *60*(4), 946–964. https://doi.org/10.2307/1131035

Bekkering, H., Bruijn, E. R. A. de, Cuijpers, R. H., Newman-Norlund, R., van Schie, H. T., & Meulenbroek, R. (2009). Joint action: Neurocognitive mechanisms supporting human interaction. *Topics in Cognitive Science*, *1*(2), 340–352. https://doi.org/10.1111/j.1756-8765.2009.01023.x

Bello, A., Sparaci, L., Stefanini, S., Boria, S., Volterra, V., & Rizzolatti, G. (2014). A developmental study on children's capacity to ascribe goals and intentions to others. *Developmental Psychology*, *50*(2), 504–513. https://doi.org/10.1037/a0033375

Brownell, C. A., & Carriger, M. S. (1990). Changes in cooperation and self-other differentiation during the second year. *Child Development*, *61*(4), 1164–1174.

Cannon, E. N., & Woodward, A. L. (2012). Infants generate goal-based action predictions. *Developmental Science*, *15*(2), 292–298. https://doi.org/10.1111/j.1467-7687.2011.01127.x

Carpendale, J. I. M., & Lewis, C. (2004). Constructing an understanding of mind: The development of children's social understanding within social interaction. *Behavioral and Brain Sciences*, *27*(1), 79-96. https://doi.org/10.1017/s0140525x04000032

Chen, Y., Keen, R., Rosander, K., & Hofsten, C. von (2010). Movement planning reflects skill level and age changes in toddlers. *Child Development*, *81*(6), 1846–1858. https://doi.org/10.1111/j.1467-8624.2010.01514.x.

Claxton, L. J., Keen, R., & McCarty, M. E. (2003). Evidence of motor planning in infant reaching behavior. *Psychological Science*, *14*(4), 354–356. https://doi.org/10.1111/1467-9280.24421

Daum, M. M., Attig, M., Gunawan, R., Prinz, W., & Gredebäck, G. (2012). Actions seen through babies' eyes: A dissociation between looking time and predictive gaze. *Frontiers in Psychology*, *3*, 370. https://doi.org/10.3389/fpsyg.2012.00370

Daum, M. M., Prinz, W., & Aschersleben, G. (2011). Perception and production of object-related grasping in 6-month-olds. *Journal of Experimental Child Psychology*, *108*(4), 810–818. https://doi.org/10.1016/j.jecp.2010.10.003

Falck-Ytter, T., Gredebäck, G., & Hofsten, C. von (2006). Infants predict other people's action goals. *Nature Neuroscience*, *9*(7), 878–879. https://doi.org/10.1038/nn1729

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Filippi, C. A., & Woodward, A. L. (2016). Action experience changes attention to kinematic cues. *Frontiers in Psychology*, *7*, 19. https://doi.org/10.3389/fpsyg.2016.00019

Flanagan, J. R., & Johansson, R. S. (2003). Action plans used in action observation. *Nature*, *424*, 769–771. https://doi.org/10.1038/nature01861

Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences*, *8*(9), 396–403. https://doi.org/10.1016/j.tics.2004.07.002

Ganglmayer, K., Attig, M., Daum, M. M., & Paulus, M. (2019). Infants' perception of goal-directed actions: A multi-lab replication reveals that infants anticipate paths and not goals. *Infant Behavior & Development*, *57*, 101340. https://doi.org/10.1016/j.infbeh.2019.101340

Ganglmayer, K., Schuwerk, T., Sodian, B., & Paulus, M. (2019). Do children and adults with autism spectrum condition anticipate others’ actions as goal-directed? A predictive coding perspective. *Journal of Autism and Developmental Disorders.* Advance online publication. https://doi.org/10.1007/s10803-019-03964-8

Giese, M. A., & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews. Neuroscience*, *4*(3), 179–192. https://doi.org/10.1038/nrn1057

Goldman, A. I. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading*. *Philosophy of mind*. Oxford University Press.

Gordon, R. M. (1995). Simulation without introspection inference from me to you. In M. Davies & T. Stone (Eds.), *Readings in mind and language: Vol. 4. Mental simulation: Evaluations and applications* (1st ed., pp. 53–67). Blackwell.

Gottwald, J. M., Bortoli Vizioli, A. de, Lindskog, M., Nyström, P., L Ekberg, T., Hofsten, C. von, & Gredebäck, G. (2017). Infants prospectively control reaching based on the difficulty of future actions: To what extent can infants' multiple-step actions be explained by Fitts' law? *Developmental Psychology*, *53*(1), 4–12. https://doi.org/10.1037/dev0000212

Gottwald, J. M., Gredebäck, G., & Lindskog, M. (2019). Two-step actions in infancy-the TWAIN model. *Experimental Brain Research*, *237*(10), 2495–2503. https://doi.org/10.1007/s00221-019-05604-0

Gredebäck, G., Stasiewicz, D., Falck-Ytter, T., Hofsten, C. von, & Rosander, K. (2009). Action type and goal type modulate goal-directed gaze shifts in 14-month-old infants. *Developmental Psychology*, *45*(4), 1190–1194. https://doi.org/10.1037/a0015667

Grosse Wiesmann, C., Friederici, A. D., Singer, T., & Steinbeis, N. (2017). Implicit and explicit false belief development in preschool children. *Developmental Science*, *20*(5), e12445. https://doi.org/10.1111/desc.12445

Harris, C. M. (1989). The ethology of saccades: A non-cognitive model. *Biological Cybernetics*, *60*(6), 401–410. https://doi.org/10.1007/BF00204695

Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision*, *3*(1), 49–63. https://doi.org/10.1167/3.1.6

Hofsten, C. von, & Rönnqvist, L. (1988). Preparation for grasping an object: A developmental study. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(4), 610–621. https://doi.org/10.1037//0096-1523.14.4.610

Hunnius, S., & Bekkering, H. (2010). The early development of object knowledge: A study of infants' visual anticipations during action observation. *Developmental Psychology*, *46*(2), 446–454. https://doi.org/10.1037/a0016543

Hunnius, S., & Bekkering, H. (2014). What are you doing? How active and observational experience shape infants' action understanding. *Philosophical Transactions of the Royal Society B*, *369*(1644), 20130490. https://doi.org/10.1098/rstb.2013.0490

Kanakogi, Y., & Itakura, S. (2011). Developmental correspondence between action prediction and motor ability in early infancy. *Nature Communications*, *2*, 341. https://doi.org/10.1038/ncomms1342

Kayhan, E., Heil, L., Kwisthout, J., van Rooij, I., Hunnius, S., & Bekkering, H. (2019). Young children integrate current observations, priors and agent information to predict others' actions. *PloS ONE*, *14*(5), e0200976. https://doi.org/10.1371/journal.pone.0200976

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35-B42. https://doi.org/10.1016/S0010-0277(02)00004-5

Kirkham, N. Z., Slemmer, J. A., Richardson, D. C., & Johnson, S. P. (2007). Location, location, location: Development of spatiotemporal sequence learning in infancy. *Child Development*, *78*(5), 1559–1571. https://doi.org/10.1111/j.1467-8624.2007.01083.x

Kochukhova, O., & Gredebäck, G. (2010). Preverbal infants anticipate that food will be brought to the mouth: An eye tracking study of manual feeding and flying spoons. *Child Development*, *81*(6), 1729–1738. https://doi.org/10.1111/j.1467-8624.2010.01506.x

Land, M. F. (2009). Vision, eye movements, and natural behavior. *Visual Neuroscience*, *26*(1), 51–62. https://doi.org/10.1017/S0952523808080899

Low, J., & Perner, J. (2012). Implicit and explicit theory of mind: State of the art. *British Journal of Developmental Psychology*, *30*(1), 1–13. https://doi.org/10.1111/j.2044-835X.2011.02074.x

Meyer, M., van der Wel, R. P. R. D., & Hunnius, S. (2016). Planning my actions to accommodate yours: Joint action development during early childhood. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *371*(1693). https://doi.org/10.1098/rstb.2015.0371

Monroy, C. D., Chen, C.-H., Houston, D., & Yu, C. (2020). Action prediction during real-time parent-infant interactions. *Developmental Science*, e13042. https://doi.org/10.1111/desc.13042

Monroy, C. D., Gerson, S. A., & Hunnius, S. (2017). Toddlers' action prediction: Statistical learning of continuous action sequences. *Journal of Experimental Child Psychology*, *157*, 14–28. https://doi.org/10.1016/j.jecp.2016.12.004

Monroy, C. D., Gerson, S. A., & Hunnius, S. (2018). Translating visual information into action predictions: Statistical learning in action and nonaction contexts. *Memory & Cognition*, *46*(4), 600–613. https://doi.org/10.3758/s13421-018-0788-6

Monroy, C. D., Meyer, M., Gerson, S., & Hunnius, S. (2017). Statistical learning in social action contexts. *PloS ONE*, *12*(5), e0177261. https://doi.org/10.1371/journal.pone.0177261

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science (New York, N.Y.)*, *308*(5719), 255–258. https://doi.org/10.1126/science.1107621

Paulus, M. (2012). Action mirroring and action understanding: An ideomotor and attentional account. *Psychological Research*, *76*(6), 760–767. https://doi.org/10.1007/s00426-011-0385-9

Paulus, M., Hunnius, S., van Wijngaarden, C., Vrins, S., van Rooij, I., & Bekkering, H. (2011). The role of frequency information and teleological reasoning in infants' and adults' action prediction. *Developmental Psychology*, *47*(4), 976–983. https://doi.org/10.1037/a0023785

Paulus, M., Schuwerk, T., Sodian, B., & Ganglmayer, K. (2017). Children's and adults' use of verbal information to visually anticipate others' actions: A study on explicit and implicit social-cognitive processing. *Cognition*, *160*, 145–152. https://doi.org/10.1016/j.cognition.2016.12.013

Pomiechowska, B., & Csibra, G. (2017). Motor activation during action perception depends on action interpretation. *Neuropsychologia*, *105*, 84–91. https://doi.org/10.1016/j.neuropsychologia.2017.01.032

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews. Neuroscience*, *2*(9), 661–670. https://doi.org/10.1038/35090060

Rizzolatti, G., & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: Interpretations and misinterpretations. *Nature Reviews. Neuroscience*, *11*(4), 264–274. https://doi.org/10.1038/nrn2805

Rosander, K., & Hofsten, C. von (2011). Predictive gaze shifts elicited during observed and performed actions in 10-month-old infants and adults. *Neuropsychologia*, *49*(10), 2911–2917. https://doi.org/10.1016/j.neuropsychologia.2011.06.018

Rotman, G., Troje, N. F., Johansson, R. S., & Flanagan, J. R. (2006). Eye movements when observing predictable and unpredictable actions. *Journal of Neurophysiology*, *96*(3), 1358–1369. https://doi.org/10.1152/jn.00227.2006.

Ruffman, T. (2014). To belief or not belief: Children's theory of mind. *Developmental Review*, *34*(3), 265–293. https://doi.org/10.1016/j.dr.2014.04.001

Ruffman, T., Taumoepeau, M., & Perkins, C. (2012). Statistical learning as a basis for social understanding in children. *British Journal of Developmental Psychology*, *30*(1), 87–104. https://doi.org/10.1111/j.2044-835X.2011.02045.x

Saffran, J. R., & Kirkham, N. Z. (2018). Infant statistical learning. *Annual Review of Psychology*, *69*, 181–203. https://doi.org/10.1146/annurev-psych-122216-011805

Schuwerk, T., & Paulus, M. (2016). Preschoolers, adolescents, and adults visually anticipate an agent's efficient action; but only after having observed it frequently. *Quarterly Journal of Experimental Psychology*, *69*(4), 800–816. https://doi.org/10.1080/17470218.2015.1061028

Sebanz, N., & Knoblich, G. (2009). Prediction in joint action: What, when, and where. *Topics in Cognitive Science*, *1*(2), 353–367. https://doi.org/10.1111/j.1756-8765.2009.01024.x

Senju, A., Southgate, V., White, S., & Frith, U. (2009). Mindblind eyes: An absence of spontaneous theory of mind in Asperger syndrome. *Science (New York, N.Y.)*, *325*(5942), 883–885. https://doi.org/10.1126/science.1176170

Smith, L. B., Jayaraman, S., Clerkin, E., & Yu, C. (2018). The developing infant creates a curriculum for statistical learning. *Trends in Cognitive Sciences*, *22*(4), 325–336. https://doi.org/10.1016/j.tics.2018.02.004

Southgate, V. (2013). Do infants provide evidence that the mirror system is involved in action understanding? *Consciousness and Cognition*, *22*(3), 1114–1121. https://doi.org/10.1016/j.concog.2013.04.008

Southgate, V., Johnson, M. H., Osborne, T., & Csibra, G. (2009). Predictive motor activation during action observation in human infants. *Biology Letters*, *5*(6), 769–772. https://doi.org/10.1098/rsbl.2009.0474

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, *18*(7), 587–592. https://doi.org/10.1111/j.1467-9280.2007.01944.x

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, *8*(3), 220–247. https://doi.org/10.1207/s15327957pspr0803_1

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., & Woolley, J. D. (1990). From simple desires to ordinary beliefs: The early development of everyday psychology. *Cognition*, *35*(3), 245–275. https://doi.org/10.1016/0010-0277(90)90024-E

Zaal, F. T. J. M., & Thelen, E. (2005). The developmental roots of the speed-accuracy trade-off. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(6), 1266–1273. https://doi.org/10.1037/0096-1523.31.6.1266

Supplementary material

Frequency of anticipations in the familiarization trials

Table 5. *Frequency of anticipations to the single cube in the familiarization phase for each age group and trial.*

Age groups	**1st trial**	**2nd trial**	**3rd trial**	**4th trial**	**5th trial**	**6th trial**
3-year-olds	0	64	46	50	29	32
4-year-olds	0	75	55	90	55	35
10-year-olds	10	50	63	83	53	60
Adults	15	85	60	75	79	65

Note: First fixations to the cube in percentages.

Frequency of anticipations in the test trials

Table 6. *Frequency of anticipations to one of the two cubes for each age group and test trial, listed for the two categories.*

Condition	**Age groups**	**1st trial**	**2nd trial**	**3rd trial**	**4th trial**	**5th trial**
Close	3-year-olds	75	57	43	46	39
	4-year-olds	75	45	65	45	50
	10-year-olds	80	87	60	70	63
	Adults	75	70	85	70	70
Far	3-year-olds	54	54	39	39	18
	4-year-olds	75	60	55	55	35
	10-year-olds	90	63	80	73	70
	Adults	95	95	85	95	70

Note: First fixations in percentages.

DLS over the five test trials.

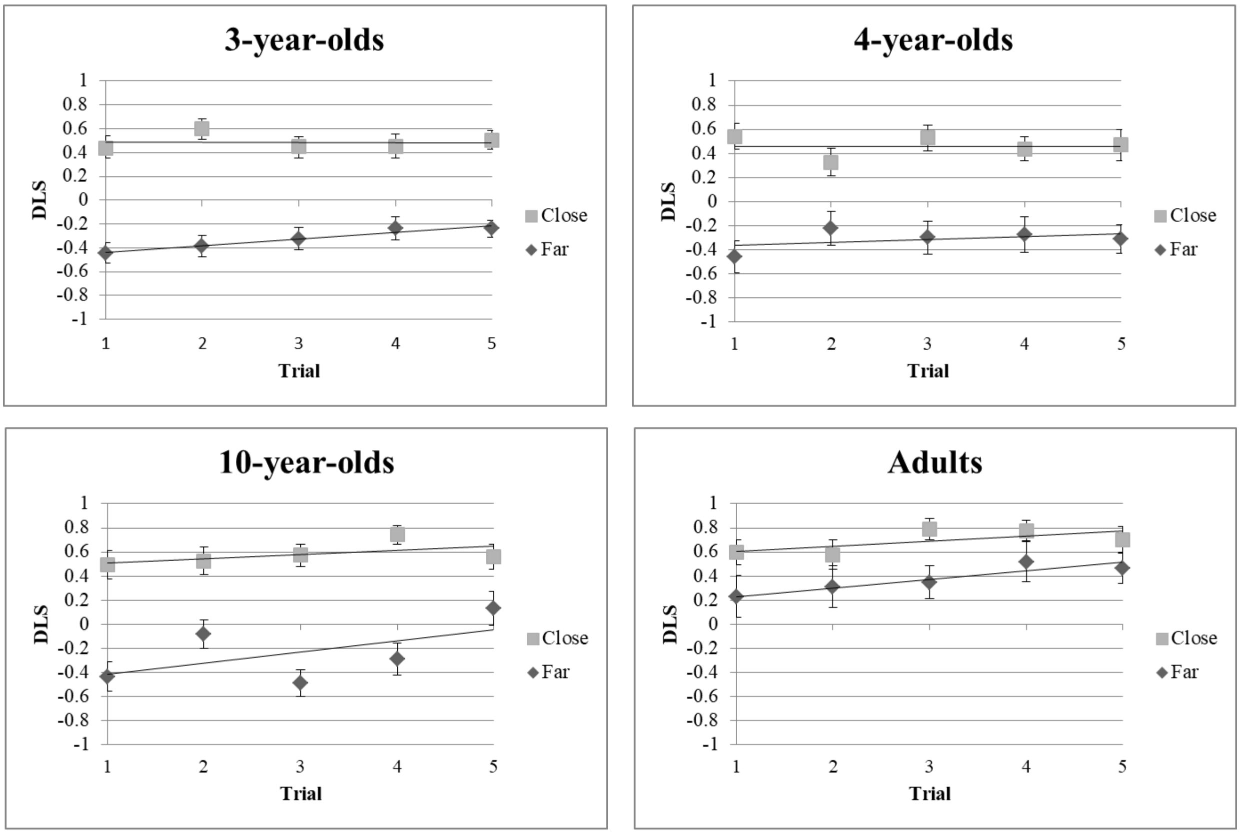


Figure 6. Differential Looking Score (DLS) over the five test trials for each age group and condition. Error bars represent standard errors. Lines depict the results of the Linear Mixed Models.

Prediction Scores over the five test trials for each condition.

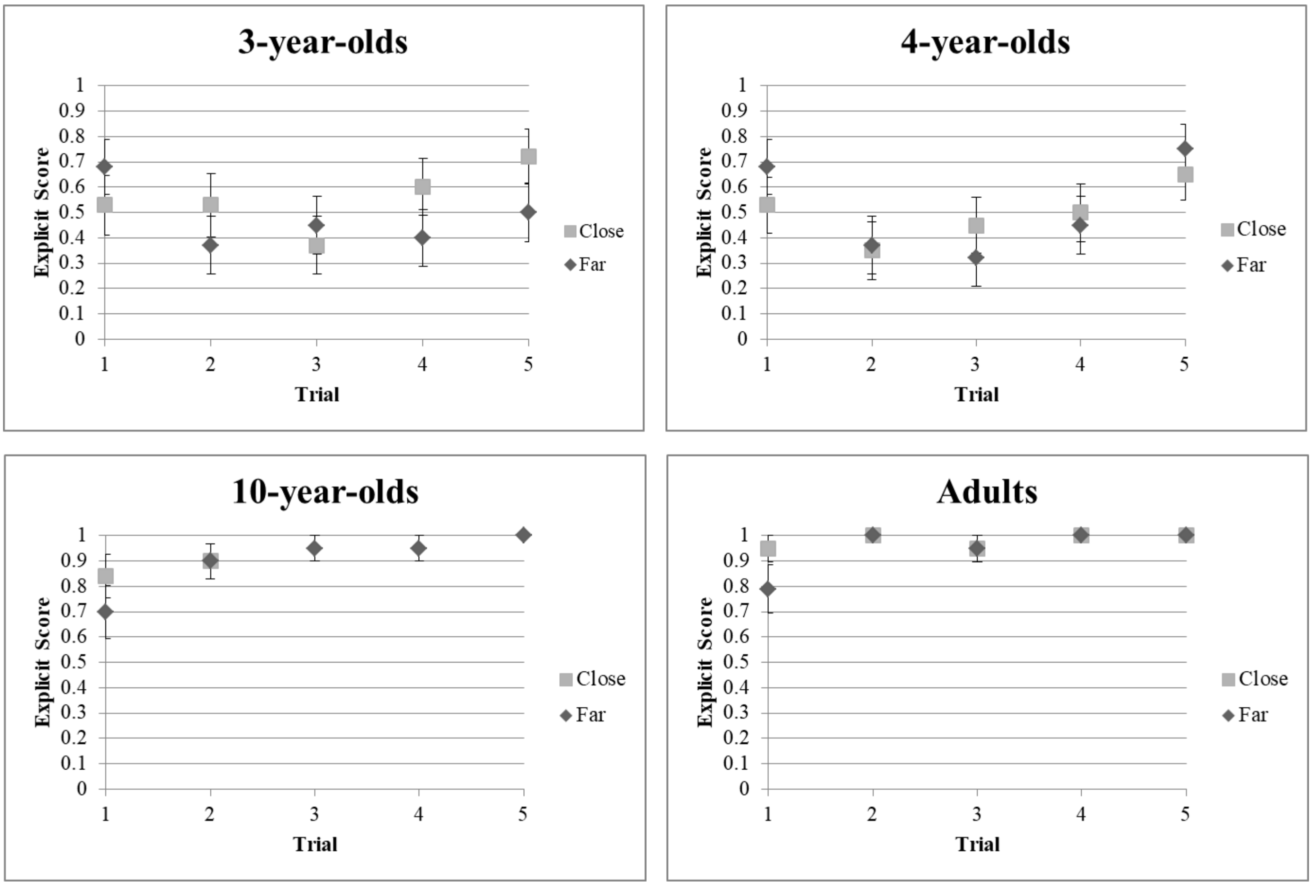


Figure 7. Explicit Prediction Scores over the five trials for each age group and condition. Chance level at 0.5. Error bars depict standard errors.

B) Study 2: The ontogeny of efficient second-order action planning: The developing interplay of controlled and habitual processes in goal-directed actions

Melzel, S., & Paulus, M. (2022). The ontogeny of efficient second-order action planning: The developing interplay of controlled and habitual processes in goal-directed actions. *Journal of Experimental Child Psychology*, *216*, 105339. https://doi.org/10.1016/j.jecp.2021.105339

C) Study 3a: The development of the prediction of complex actions in early childhood

Melzel, S., & Paulus, M. (2021). The development of the prediction of complex actions in early childhood. *European Journal of Developmental Psychology*, *18*(2), 161–183. https://doi.org/10.1080/17405629.2020.1773786

D) Study 3b: The development of processing of nested structures in preschool children

Melzel, S., & Paulus, M. (submitted). The development of processing of nested structures in preschool children.

Abstract

Influential theories have proposed that processing nested structures constitutes an important characteristic of human cognition among various domains and might represent an important cognitive capacity in human development. The aim of the current study was to investigate how the ability to process nested structures develops in early childhood. Therefore we assessed the development of nested structure processing (NSP) in three- to six-year-old German children (N = 130) across two domains, language and action. We explored whether NSP development is related across the domains, to which extent it is driven by underlying general cognitive functions (inhibitory control, working memory) and whether it contributes to the early development of Theory of Mind (ToM). We found NSP in the action and language domain to correlate. Whereas the correlation remained when controlling for inhibitory control, it was not significant anymore when controlling for working memory. Furthermore, we found NSP to predict ToM and to explain age effects in ToM even when controlling for working memory and inhibitory control. Overall, the findings point to the development of a domain-general NSP capacity during preschool years. They suggest, though, that working memory constitutes an essential basis for NSP. Furthermore, our results support cognitive theories proposing that NSP plays an important role for the early development of an explicit ToM.

Keywords: hierarchy, preschoolers, syntax, action, Theory of Mind, recursion

The development of processing of nested structures in preschool children

The processing of nested structures (NSP) has been proposed to constitute an important characteristic of human cognition (e.g., Corballis, 2014; Dawkins, 1976; Lashley, 1951; G. A. Miller et al., 1960). In nested structures, elements are arranged hierarchically so that higher level elements span over lower level units and in turn lower level units are nested into the higher level ones (Botvinick, 2008; Fengler et al., 2016; G. A. Miller et al., 1960). NSP seems to be largely unique to humans, showing limited capacities in great apes and chimpanzees (Byrne & Russon, 1998; Conway & Christiansen, 2001; Ferrigno et al., 2020; Fitch & Hauser, 2004; Greenfield, 1991). Influential theories have proposed that nested structures play a crucial role among various domains of human cognitive functioning. Moreover, it has been suggested that NSP might constitute a domain-general capacity (Fadiga et al., 2009; Grossman, 1980; Jeon, 2014; Lashley, 1951; Marcus, 2006). Developmental theories have proposed that the processing of hierarchical structures might constitute an important capacity underlying cognitive development (e.g., Corballis, 2014; Greenfield, 1991; Martins et al., 2014).

More specifically, in the language domain it has been claimed to be pivotal to understand complex sentences (e.g., Fengler et al., 2016; Hauser et al., 2002). Relatedly, concerning action and behavioural control it has been suggested to be important to guide behaviour according to higher level goals which allows for more complex and efficient actions. For example, hierarchically structured actions—opposed to linear ones—can be repaired more easily when they fail (Dawkins, 1976; G. A. Miller et al., 1960) and are central for human tool-use (Byrne & Russon, 1998; Gönül et al., 2018). In the domain of social cognition it has been claimed to be relevant to understand that someone else’s state of knowledge might differ from one’s own, as what is believed to be someone else’s state of mind has to be nested into one's own state of mind (Corballis, 2014; Frye et al., 1995; Ryle & Tanney, 2009). Thus, the processing of nested structures seems to play a key role in different forms of human higher-order cognition.

Given these long-standing theoretical claims on the role and the domain-general nature of hierarchical processing (Fadiga et al., 2009; Jeon, 2014; Lashley, 1951) it is surprising that little is known about how the ability to process hierarchical structures develops in early childhood. This study investigated the early development of processing nested structures across two different domains, language and action. It explored whether NSP in different domains is related to each other, to which extent it is driven by underlying general cognitive functions (inhibitory control, working memory), and whether NSP makes a contribution to Theory of Mind (ToM). It aims at contributing to theoretical debates on the extent to which nested processing is a domain-general process.

The development of nested structures in various domains

It is widely accepted that language—and in particular syntax as a subdomain of grammar—shows a hierarchical structure (Chomsky, 1956; Friederici et al., 2011; Greenfield, 1991). Combining several words (units) into phrases and sentences (higher-order element) results in syntactic hierarchies. Nested structures, resulting from grammars of the form A^n B^n and showing nested relations of A and B (A_1 (A_2 (A_3 B_3) B_2) B_1), constitute an important type of syntactic hierarchies. They are essential to human language, as processing such structures allows for understanding nested sentences such as nested relative clauses ("The cat [that is black] is chasing the mouse."). Note that the nesting can be repeated infinitely, reflecting the recursive structure of language ("The cat [that the dog [that is tall] is chasing] chases the mouse.").

To understand such structures it is essential to process and maintain the thematic relationships of the noun phrases, i.e., identifying who is doing what to whom. In centre-nested relative clauses (as in the examples above) the long-distance dependencies constitute a major challenge for children, as they need to inhibit their preferential interpretation strategy of relying on the word order (Fengler et al., 2015; Lindner, 2003). Previous studies found significant improvements in the processing of centre-nested relative clauses between three and five years of age (e.g., Fengler et al., 2016; Kidd & Bavin, 2002; Villiers et al., 1979) and hint to further improvements during middle childhood when processing more complex double-nested relative clauses (Fengler et al., 2016). The findings are of vital importance with respect to the development of NSP within the language domain. However, they leave the question open whether NSP develops as domain-general capacity.

It has been suggested that not only language but also complex behaviour shows a nested structure: higher level goals control sub-goals which in turn control more simple actions (Botvinick, 2008; Cooper et al., 2014; Dawkins, 1976; Lashley, 1951; G. A. Miller et al., 1960). According to Dawkins (1976) one advantage of hierarchically over linearly organized behaviour is that the former is easier to repair when it fails than the latter. He refers to an example (introduced by Simon, 1962) of two watchmakers: One watchmaker tries to put together all 1000 components of a watch sequentially in one single assembly process. Another watchmaker shows a hierarchical, nested approach: To complete a watch (goal) he builds and combines 10 large sub-assemblies (sub-goal). To build one of the 10 large sub-assemblies he builds and combines 10 medium sub-assemblies (sub-sub-goal). To build in turn one medium sub-assembly he builds and combines 10 small sub-assemblies consisting of 10 components each. In case of failure the first watchmaker might have to start from scratch whereas the second watchmaker would only need to reassemble the defective subunit. Note that the second watchmaker's goal of assembling the watch spans over all nested sub-goals and actions. Such a nesting of sub-goals can be repeated infinitely, illustrating the recursive structure of behaviour and, thus, the comparability to the language domain.

Whereas pre-schoolers are able to imitate and learn novel sequential actions (e.g., Buchsbaum et al., 2011; Yanaoka & Saito, 2019) their ability to process nested structured actions significantly improves during preschool years (Bello et al., 2014; Flynn & Whiten, 2008; Freier et al., 2015, 2017; Greenfield & Schneider, 1977; Melzel & Paulus, 2021). For example, in a study by Freier et al. (2017) children were instructed to colour in six shapes according to an overarching goal (using each of three crayons equally often). Whereas both, 3- and 5-year-olds showed good performance at the lowest level of the action hierarchy (colouring-in), only 5-year-olds consistently aligned their colouring-in action to the higher level goal of using each crayon equally often. The findings might be explained with respect to young children's hierarchical goal representation being too weak to switch between various goals so that they rather rely on their procedural knowledge (colouring-in activity). This is also in line with findings that 3-year-olds can distinguish and name simple actions such as touching and grasping when observing the respective motor act. However, only 6- to 7-year-olds are able to discern the goal-related—hierarchically higher—motivations behind the motor-acts (e.g., an object is being grasped to be used vs. to be placed) (Bello et al., 2014). Taken together previous research points to significant improvements in NSP in the action domain during preschool years and, thus, shows parallels to the developmental pathway of NSP in the language domain. Though, until today it has not been examined, whether the development of NSP is related among these two domains.

Another domain in which nested structures have been proposed to play a role is Theory of Mind (ToM), that is, the attribution of mental states to oneself or other persons (Astington & Jenkins, 1999; Premack & Woodruff, 1978). ToM is often assessed using false-belief tasks, which require explicit reasoning about another person's knowledge state. In these tasks the belief of another person conflicts with one's own knowledge and with reality. In a classical false-belief task "Paul thinks his mittens are in the wardrobe". Actually, though, his mittens are not in the wardrobe but in his backpack. Hence, there is a discrepancy between reality and Paul's belief. Frye et al. (1995) proposed that children's success in ToM tasks is based on their ability to apply hierarchical, embedded—i.e., nested—rules. They proposed that, what they called, hierarchically higher setting conditions are needed to encode the perspectives of self and other and that the judgements are nested under the respective perspective. Returning to the example of Paul's mittens, they suggested that only if children can switch between the hierarchically higher setting condition of "own knowledge / reality" and "other's belief" they are able to solve false-belief tasks. In the same vein, Corballis (2014, p. 133) hypothesized that ToM "is recursive, in the sense that it involves the insertion of what you believe to be someone else's state of mind into your own". The recursive nesting becomes obvious in second-order false belief tasks in which children have to appreciate that another person may have a false belief about someone else's (true or false) belief (e.g., S. A. Miller, 2009). The nesting can be repeated infinitely as in the action and language domain, pointing to parallels among the three domains.

Understanding explicit false-belief tasks significantly improves around 4 to 5 years of age (e.g., Astington & Jenkins, 1999; Flavell, 2004; for review see Wellman et al., 2001). This points to developmental changes at the same age as significant improvements in NSP in the language and action domain can be found and, thus, might hint to similar underlying processes. A few studies have started to investigate the relevance of hierarchical structure processing in the emerging ToM (Frye et al., 1995; Grosse Wiesmann, Schreiber, et al., 2017; Villiers & Villiers, 2000). Frye et al. (1995) for example designed two non-mental-state tasks (card-sorting and causality task) that required applying nested rules. They found three- to five-year-olds' performance in these tasks to correlate with ToM, suggesting that applying nested rules account for developmental changes in ToM. Additional evidence comes from a neuroimaging study indicating that the neural basis of ToM overlaps with the neural correlates of hierarchical processing (Grosse Wiesmann, Schreiber, et al., 2017). In particular, this study reported that developmental changes in the arcuate fasciculus (more specifically in the inferior frontal gyrus (IFG), an area supposed to be involved in processing hierarchies) correlate with 3- and 4-year olds' ToM ability. Given these considerations and findings NSP might also be relevant for (the emerging) ToM. The current study will add to this line of research and examine the contribution of NSP in the action and language domain to the emerging ToM.

Processing nested structures—a domain-general capacity?

Even though the structures in each of the three domains show domain-specific characteristics it has been proposed that they are comparable among the different domains (Fadiga et al., 2009; Maffongelli et al., 2019; Marcus, 2006; Pulvermüller & Fadiga, 2010). In all domains hierarchically lower units are nested into hierarchically higher elements, so that the hierarchically higher elements span over the lower units. Thus, they cannot be analysed as sequential Markov chain. Returning to the examples provided above: in the language domain "The cat is chasing the mouse" spans over the nested relative clause "that is black", in the action domain the goal of assembling the watch spans over the sub-goals of building sub-assemblies, and in ToM the attribution of a belief spans over the nested content of that belief. These kinds of nesting structures have also been referred to as $A^n B^n$ Grammar (Friederici et al., 2011), "correct nesting of brackets" (Dawkins, 1976), or "higher order actions" (Ryle & Tanney, 2009) and are characterized by recursion (theoretically infinitely repetition of nesting). Given the parallels among the domains, NSP might be a multi-domain capacity of human cognition (Fadiga et al., 2009; Jeon, 2014; Koelsch et al., 2013; Lashley, 1951).

Assuming that there is a parallel development of processing nested structures across different domains, an intriguing question concerns whether some underlying basic cognitive processes can account for it. According to information processing theories higher cognitive processes such as problem solving or reasoning rely on simpler mechanisms such a working memory (WM), processing speed, inhibitory control (IC), and the acquisition of more efficient strategies and greater knowledge (e.g., Kail et al., 2016; Richland & Burchinal, 2013; Zheng et al., 2011). As WM and IC are crucial to maintain task and goal-related information while inhibiting task-irrelevant information one could hypothesize that they might be especially important for the development of processing nested structures. The ability to maintain hierarchically higher-level information while processing hierarchically lower-level information as well as to inhibit predominant responses seems important in all three domains: (a) in the language domain the noun phrase of the main clause has to be maintained across the nested clause. Furthermore, the premature release of the information of the nested clause has to be suppressed to process the main clause; (b) in the action domain the overarching goal has to be maintained while processing and executing the actions of lower-level sub-goals. Predominant responses as for example continuing low-level actions without aligning them to higher-order goals have to be inhibited; (c) in false-belief paradigms the attribution of a belief to a person has to be maintained while processing the belief itself and while inhibiting one’s own predominant perspective. Interestingly, developmental studies within these domains have shown that WM and IC relate to processing nested structures in language (e.g., Fengler et al., 2016; Ferrigno et al., 2020; Meyer et al., 2013) as well as to false-belief understanding (Devine & Hughes, 2014; Grosse Wiesmann, Friederici, et al., 2017; Milligan et al., 2007; Moses, 2005). Thus, if NSP turns out to be related across different domains, it would be interesting to see whether more basic domain-general cognitive abilities such as WM and IC can account for this relation.

Goal of the current study

The current study aims at targeting the nature of nested structures and its role for three domains of early cognitive development. More precisely, it has three main aims: (1) To examine whether the early development of NSP is related across the action and language domain. If it develops as domain-general capacity in line with previous theoretical considerations (Corballis, 2014; Fadiga et al., 2009; Gönül et al., 2018; Greenfield, 1991; Jeon, 2014) we would expect a related development across domains. (2) To investigate to which extent potential developmental parallels in NSP across the domains can be explained by general cognitive abilities such as WM or IC. There seem three possible interrelations: (a) if NSP is independent of WM and IC the relation across domains should persist when controlling for WM/IC; (b) if NSP can be explained by WM/IC potential relations across the domains should vanish when controlling for it; and (c) if WM/IC constitute the basis for NSP but NSP develops as domain-general capacity beyond WM/IC potential relations across the domains would be reduced but remain when controlling for it. (3) Whether processing nested structures in the language and action domain predicts early ToM beyond other domain-general functions. If so, it would point to the importance of NSP for the development of ToM.

In order to address the questions we assessed three- to six-year old children. We administered two language domain tasks (single nested and double nested relative sentence comprehension task), three action domain tasks, two false-belief tasks (content and explicit false-belief), a WM task (digit-span forward) and an IC task (Day-and-Night).

Method

Transparency and openness

We report how we determined our sample size, how we dealt with missing data, all manipulations and all measures in the study, and we follow JARS (Kazak, 2018). The data that support the findings of this study as well as the analysis code are openly available in the

Open Science Framework (OSF) at https://osf.io/g3ftx/?view_only=c9e2e9da9b2744b49986af9b0712d95b (anonymous view-only link, will be made public after acceptance). Data were analysed using R, version 3.6.1. This study's design and its analysis were not pre-registered.

Participants

One-hundred-thirty 3 to 6 year old monolingual German children (64 females, age range: 41-79 months, *M* = 59.48) were included in the final sample. All children were born after the 37th week of pregnancy. Children came from a large city in Germany and were recruited from birth records. Informed consent for participation was given by the children's caregivers. Parents received 5€ travel compensation and children got a small present. Ethic approval was obtained from the local ethics board and the treatment of the sample complied with APA ethical standards.

Power and sample size

As previous studies within the action domain showed medium to high effect sizes when comparing 3- to 5-year-olds' performances in NSP (Freier et al., 2015, 2017; Melzel & Paulus, 2021) we decided to test for a medium effect size when correlating the two domains. An a priori power analysis using a two tailed bivariate normal model with a presumed medium effect size of $r = 0.3$, $\alpha = 0.05$ and power = 0.80 with G*Power 3.1.9.2 (Faul et al., 2009) yielded a sample size of at least 84 participants. Moreover, previous studies point to strong relations between the developing syntactic processing and ToM (Grosse Wiesmann, Friederici, et al., 2017). Hence, we decided to run an a priori power analysis for a hierarchical, linear, multiple regression with a presumed medium effect size of $r = 0.3$, $\alpha = 0.05$, power = 0.80, two tested predictors (language and action) and four predictors in total (language, action, WM, IC) with G*Power 3.1.9.2 (Faul et al., 2009). As it yielded a sample size of 101, we aimed at assessing more than 101 children.

Overview of the tasks.

To assess NSP in the language domain, we applied two sentence comprehension tasks, one using single and one double nested relative sentences. Three tasks assessed NSP in the action domain: the means-end action prediction, colouring-in, and tree task. Furthermore, we assessed general cognitive abilities with two tasks (WM: digit span, IC: day&night) and ToM with two false-belief tasks (content and explicit false-belief).

Materials, Setting and Procedure

The study was conducted in the laboratories of the LMU Munich. Each participant was tested on all nine tasks. For reasons of statistical comparability in a correlational approach the sequence of tasks (means-end prediction, single and double nested language, tree, digit span, content false-belief, explicit false-belief, day&night, colouring-in) as well as the item presentation sequences within each task were kept constant. Participants sat across the experimenter and materials were presented on a table. Overall, a test session lasted approximately 50 minutes.

Language tasks. Material was taken from Fengler et al. (2016). Each sentence consisted of three clauses that were arranged according to two levels of complexity: in the single nested task two coordinated relative clauses were nested, in the double nested task one relative clause was nested in the other and the third was again nested in the already nested clause (see Figure 1). For more details on item construction see Fengler et al. (2016). Picture pairs of an incorrect and a correct picture were assigned to each sentence. For each task, six picture pairs tested for the comprehension of the long-distance dependency between the initial subject (e.g., “The bird”) and the final verb of a sentence (e.g., “is blue”). In two additional picture pairs the crucial information was part of the nested clause, to prevent participants from the strategy of only processing the first and the last part of the sentence. Additionally, three picture pairs of known objects (e.g., apple, banana) and simple sentences referring to one of the two pictures (e.g., “This is an apple”) were used for familiarization purposes.

Sentences and pictures were presented using Presentation® software (Version 18.0, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com). In both, familiarization and testing trials, two pictures were presented on a screen next to each other, one matching and one mismatching the auditory stimulus. The experimenter instructed the child to press the button that corresponded to the position of the image that matched the auditory stimulus. Children responded by pressing the left or right button of a Bluetooth mouse which they were holding in their hand. Participants completed three familiarization trials to ensure that they understood the task and to get used to the mouse clicking. They received feedback after each familiarization trial. The procedure was repeated if they responded incorrectly in at least one out of the three trials. Except for one child, whose data of this task was excluded from further analyses, all participants correctly responded to the three trials within three attempts. After familiarization, the experimenter explained that they were going to play the same game again but this time the TV was going to read out the sentences. Each task (single nested, double nested) consisted of eight items that were presented intermixed, hence, children were tested on 16 language items in total. The presentation of the correct image on the left or right was counterbalanced. Response time was limited to 10s and a white screen was presented between trials.

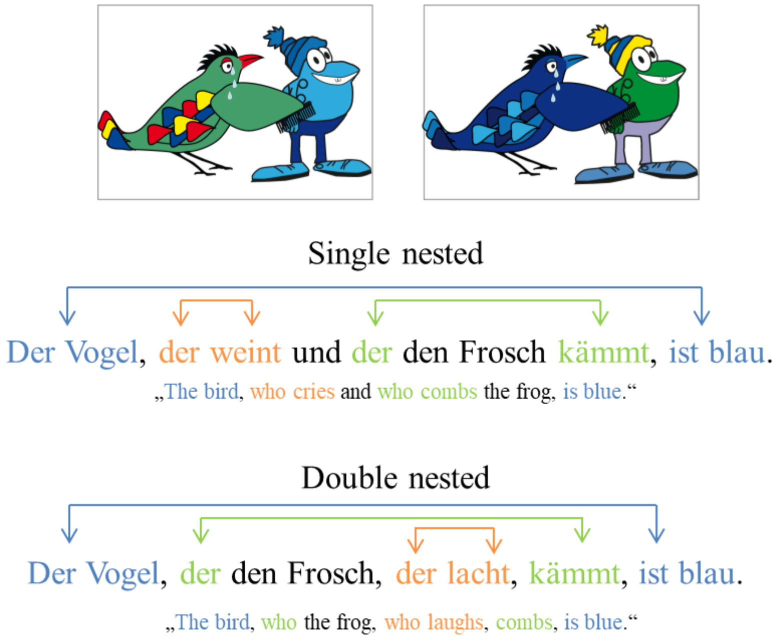


Figure 1. Language tasks. Top: example picture pair (left incorrect, right correct) of the two picture matching tasks. Middle and bottom: example sentences of the single and double nested task.

Action tasks.

Colouring-in task. Materials were adapted from the study by Freier et al. (2017). To introduce the concept of (in)equality the experimenter showed equal and unequal distributions of six plastic mice on two plates. In a pre-test participants had to indicate the type of distribution (equal, unequal) and the experimenter gave feedback. The pre-test was terminated after participants correctly responded to two consecutive trials. All participants passed the pre-test within two attempts, showing an understanding of the concept of equality. The subsequent testing phase followed the procedure of Experiment 1 by Freier et al. (2017). Instead of animals as in the original study, six geometrical shapes were depicted on an A5 paper in the testing phase (see Figure 2, Colouring-in). Geometrical shapes were used to exclude possible colour effects for specific items as they might occur for animals (e.g. colouring in all pigs in pink). Participants were instructed to colour-in the six shapes from left to right, as indicated by an arrow beneath the shapes (see Figure 2, Colouring-in).

Furthermore, they were asked to colour-in all shapes. Then they were told the overarching goal of the game, that is, to use each of the three crayons equally often. To ensure that children paid attention to the instruction they were asked to repeat the goal. If a child was not able to repeat the goal the experimenter helped the child by giving cues of increasing informational value, reaching from "How do you have to colour-in these shapes?" over "Do you have to use all colours equally often or only a single colour?" to a repetition of the goal. The experimenter showed the participant three example pictures (depicting other geometrical shapes and colours as in the testing phase) in which three colours were used equally often. Participants were then asked to colour-in their shapes. For colouring-in the experimenter randomly selected three crayons (different colours than the ones used in the sample pictures). After children had coloured-in the first shape the experimenter reminded them of the overarching goal. Apart from the reminder the experimenter pretended to do something else while children were colouring-in their shapes.

Tree task. For the tree structure task 20 small wooden sticks (5cm x 0.5cm x 0.5cm) were placed within children's reach on the table. Moreover, two laminated A4 photos displayed the goal-state trees (see Figure 2, Tree). In the easy condition the photo matched the structure introduced by Gönül et al. (2018) and in the difficult condition the structure matched the original tree-structure by Greenfield and Schneider (1977). To familiarize children with the material the experimenter demonstrated how to build a triangle, asked the child to build one as well, and praised the child for doing so. Afterwards the experimenter positioned the picture of the easy tree clearly visible in front of the child. The child was then asked to rebuild the figure using the wooden sticks. If participants indicated that they were not able to do so the experimenter encouraged them once. While children built the tree, the experimenter pretended to do something else. If children did not start within 60s the task was terminated. After participants signalled that they were done the experimenter praised them and put the sticks back to the side. The procedure was repeated with the difficult tree.

Means-end action prediction task. Material and procedure were adapted from Melzel and Paulus (2021). Six laminated A4 paper cards each showing object pairs located at the right top and bottom of a circular path were used for testing (see Figure 2, Means-end prediction). Action goals (e.g., "I want to feed the horses") were assigned to each testing card and required the actor to nest an action (means, e.g., get the food) before going to the final goal item (end, the horses). In the familiarization phase we ensured that participants would know all actions and introduced them to the path paradigm. In the testing phase the experimenter placed the paper cards one after each other on the table in front of the participant so that they were clearly visible and easily reachable for pointing. Children were (differing from the original study) told that the actor wanted to achieve her goal as fast as possible to clarify that the actor would always take the shortest path. For each card the experimenter stated: "Look, here is Marie (again) and says: ‚I want to feed the horses'. Now she starts moving [pointing along the path between the actor and the occluder, describing a forward movement] and is hidden underneath the bridge [pointing to the occluder]. And show me, where will she reappear, if she wants to feed the horses?" (underlined phrases were replaced by the respective action goal). No feedback was given during the test phase. Children were tested on six nested action goals. Whether the top or bottom path was correct was counterbalanced.

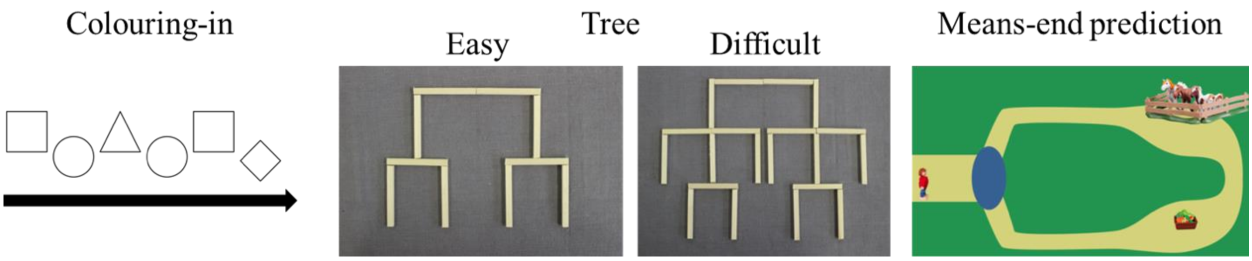


Figure 2. The three action domain tasks: (A) Colouring-in task: in this task participants had to nest the goal of colouring in the shapes into the overarching goal of using each of three crayons equally often. Without nesting the action they were assumed to just follow the low-level activity of colouring in the shapes aligning their action to the higher-level overarching action goal. (B) Tree task: the task should be considered as an analogy: being able to construct nested shapes is assumed to reflect a hierarchical cognitive organization, as it requires to not only see the parts and the whole, but to see them simultaneously and in relation to each other. (C) Means-end prediction task: To correctly predict the action participants had to nest the two sub goals (getting food and going to the horses) into the overarching goal (e.g. feeing the horses).

General cognitive abilities.

Working memory. The classical digit span forward task of the Kaufman Assessment Battery for Children (KABC-II, Kaufman & Kaufman, 2015) was carried out to assess WM. After explaining the task, the experimenter read out the number sequences which the participant was instructed to repeat. The easiest sequences consisted of two numbers and the difficulty level increased up to sequences of nine numbers. The task was terminated when children failed to correctly repeat three consecutive number sequences.

Inhibitory control. As a measure of inhibitory control we used the day&night task following the procedure of study 1 by Simpson and Riggs (2005). The material consisted of two types of laminated paper cards. A “day card” (D) depicted a yellow sun on a white background and a “night card” (N) a white moon and several stars on a black background.

The day and night cards were introduced and children were instructed to respond to the day card with “night” and to the night card with “day”. In a pre-test two pictures were presented (DN). If children responded incorrectly to at least one item the procedure was repeated up to two times. In the testing phase 14 day and night cards were presented in a pseudorandomized order (NDNDDNNDNDDNDN). In contrast to Simpson and Riggs (2005) we did not limit the presentation time of the cards, though, the experimenter continued with the next card if the child did not respond. Feedback was given only during the pre-test.

Theory of Mind. Materials for the content false-belief (CFB) and explicit false-belief (EFB) task consisted of the suggested ToM scale material: a smarties box, a small toy cow, two toy figures and a picture illustrating a wardrobe as well as a backpack (Hofer & Aschersleben, 2005; Wellman & Liu, 2004). The tasks were conducted as specified in the ToM scale by Hofer and Aschersleben (2005).

Content false-belief task. In the CFB task participants had to guess the content of a smarties box. After children’s initial guess that the box contained smarties, the experimenter showed that it actually contained a cow. The cow was put back into the box and children were asked once again: “what is in the box?” If they answered incorrectly the procedure was repeated. After children stated that a cow is in the box the experimenter introduced a toy figure, which had never looked in the box before. The child was then asked what the toy figure thinks is in the box. As a control question children had to indicate whether the toy figure had previously looked into the box.

Explicit false-belief task. In the EFB task the experimenter showed participants a toy figure (Paul) and a picture of a wardrobe and backpack. The experimenter explained that Paul was searching for his mittens that could either be in the wardrobe or in his backpack. After telling the child that Paul’s mittens are in the backpack but that Paul thinks his mittens are in the wardrobe the experimenter asked the child where Paul will search for his mittens. As a control question children were asked about the actual location of Paul's mittens.

Coding

Language scores. For each language task a response was considered correct and coded as 1 if participants pressed the corresponding button to the matching picture (left button when the matching picture was presented on the left side and right button when it was presented on the right side). Pressing the button corresponding to the non-matching picture was coded as 0. If a participant did not respond within 10s the trial was coded as "no answer given".

Language scores were calculated for each task (single nested, double nested) by dividing the sum of correct responses by the total number of given answers, leading to a score ranging from 0 to 1. An overall language score including answers of both tasks was calculated in the same way. The overall language score was used in all further analyses.

Action.

Colouring-in score. To measure participants' ability to control their action according to the overarching goal we coded how often each crayon was used. It was coded as 1 if children used each crayon exactly twice and as 0 otherwise.

Tree score. To measure participants' ability to rebuild hierarchical structures a hierarchical complexity score was built following Greenfield and Schneider (1977). The scoring method traces back to graph theory. In graph theory the junctions of lines are called nodes. The degree of a node changes depending on how many lines join. Two lines joining the resulting node was coded as 2^2, three lines as 3^2 and so on. If a line did not join any other line the end nodes were coded as 1^2. Two complexity scores were calculated by summing up the values of all nodes: one for the easy and one for the difficult tree structure. These scores ranged from 20 to 50 (using all 10 sticks) for the easier structure and from 40 to 102 (using all 20 sticks) for the difficult structure. Since all 20 sticks were on the table from the beginning, some children used more than 10 sticks to build the easy tree. If children built figures of greater complexity than the template, their score was limited to 50 or 102, respectively. If the structure was copied exactly the complexity score was raised by 10 for the easy structure (to a maximum of 60) and by 20 (to a maximum of 122) for the difficult structure. The final hierarchical complexity score was calculated by dividing the sum of the easy and difficult complexity score by the maximum achievable score (60+122 = 182). Thus, the score could range from 0.01 (1 stick used per structure) to 1 (perfect copies of both structures).

Means-end action prediction score. Participants' pointing to either of the two paths was assessed. Pointing to the correct path that is the path leading to the means was coded as 1 whereas pointing to the incorrect path was coded as 0. A means-end action prediction score was calculated by dividing the sum of correct responses by the total number of given answers.

General cognitive abilities.

Working Memory. The score of the digit span task was calculated by dividing the amount of completed sequences minus the number of incorrect sequence repetitions by the maximum possible number of correct answers (21). Thus, it ranged from 0 to 1.

Inhibitory control score. The inhibitory control score was calculated by dividing the number of correct responses by the total number of given answers.

Theory of Mind. In each ToM task participants scored a 1 if they correctly responded to both target and control question and a 0 otherwise. The scores of the two tasks were summed up to a common ToM score, ranging from 0-2.

Content false-belief. In the CFB task the correct and incorrect answers to the target question ("what does the toy figure thinks is in the box?") were "smarties" (coded as 1) and "cow" (coded as 0). The correct and incorrect answers to the control question ("has the toy figure looked in the box before?") were "no" (coded as 1) and "yes" (coded as 0).

Explicit false-belief. In the EFB task the correct answers (coded as 1) were "wardrobe" to the target question ("where will he look for his mittens?") and "backpack" to the control question ("where are his mittens in reality?"). The incorrect answers (coded as 0) on the other hand were "backpack" and "wardrobe".

Analysis strategy

To find out whether NSP develops as multi-domain capacity we decided to correlate the scores of the action and language domain. To assess whether a potential relation between the two domains was due to general cognitive abilities we planned to control for WM and IC. To find out whether the ability to process nested structures in the language and action domain predicts ToM beyond other domain-general functions we decided to run hierarchical regressions.

Results

Table 1 shows the descriptive data for each task. All analyses were run in R version 3.6.1 on the previously described scores. Multiple imputations using the package "mice" were applied to account for missing data, which were assumed to be missing at random. Five imputed datasets were created. Unless specified differently, all following analyses were performed on the five imputed datasets and estimates were pooled subsequently.

To reduce the three action scores (Colouring-in, Tree, and Means-end action prediction score) to a single nested action score we calculated an explorative Maximum-Likelihood factor analysis on the imputed datasets and pooled the results. Note that computing an explorative factor analyses (EFA) and using the resulting factor scores in further analyses is a common approach (DiStefano et al., 2009). The scores of the three tasks were suitable for a factor analysis (KMO: 0.561, Bartlett χ^2 (3) = 42.750, p < .001). The resulting factor explained 39.478% of the variance with factor loadings of .534 (Colouring-in score), .888 (Tree score), and .315 (Means-end action prediction score). Thompson's scores were calculated and used as a composed nested action score in further analyses.

The language score was calculated as the mean of the single and double nested language task. As we were nevertheless interested in whether children performed better in one of the tasks, a two-sided paired sample t-test was run on the non-imputed dataset. It showed that children performed significantly better in the single-nested (*M* = .696, *SD* = .199) compared to the double-nested language task (*M* = .588, *SD* = .204); *t*(122) = 5.480, *p*<.001. Furthermore, a correlation between the single and double nested language score turned out significant (*r* = .416, *p* <.001).

Table 1. Number of participants, means, and standard deviations for the different tasks (N = 130).

	N	**M**	**SD**	**n (score=1)**
Language				
Single nested	123	0.696	0.199	
Double nested	123	0.588	0.204	
Action				
Colouring-in	120			93
Means-end	129	0.641	0.241	
Tree	118	0.773	0.237	
General cognitive abilities				
WM	125	0.363	0.110	
IC	121	0.762	0.244	
ToM				
CFB	128			53
EFB	128			62

Does NSP develop as domain-general capacity?

One main question of the study is whether the early development of NSP is related among the action and language domain. To answer this question, a Pearson correlation between the nested action and language score was run on the imputed datasets and the pooled result turned out significant ($r = .246$, $p = .006$), pointing to similar developmental pathways in both domains.

The role of general cognitive abilities

Given that the early origins of NSP correlated among the language and action domain, the second main question was whether general cognitive abilities (WM, IC) can account for this parallel development.

To investigate whether WM and IC play a role for NSP in language and action, further correlations between general cognitive abilities and nested action and language were run on the imputed datasets. All pooled correlations got significant (see Table 2) pointing to an important role of WM and IC within the domains.

Table 2. Correlational results between general cognitive abilities and nested action and language (N=130).

General cognitive ability	Nested structure domain	*r*	*p*
WM	Language	.457	<.001
WM	Action	.451	<.001
IC	Language	.199	.026
IC	Action	.284	.005

Interestingly, although IC was related to each domain, the correlation between NSP in the language and action domain remained significant when controlling for IC ($r = .202$, $p = .028$). However, the relation between the domains disappeared when controlling for WM ($r =$

.050, p = .592), pointing to the crucial role that WM seems to play for the development of processing nested structures[1].

Nested structure processing and ToM

To investigate whether NSP in the action and language domain is predictive for ToM we ran hierarchical regressions on the imputed datasets and pooled the results (see Table 3). As WM and IC seemed relevant for NSP they were entered first in a hierarchical regression. Overall, this model was significant, $F(2,127) = 7.128$, $p = .001$, $R^2 = 0.101$, adjusted $R^2 = 0.087$. Participants who performed better in the WM task performed better in ToM. IC on the other hand was not a significant predictor of ToM. In a second step, nested language and action were added to examine the predictive value of nested structures after controlling for WM and IC. The addition of these variables was significant, $\Delta F(4, 125) = 6.202$, $p = .003$, $\Delta R^2 = .081$. Participants who performed better in NSP in the action domain showed a higher performance in ToM, whereas NSP in the language domain was not a significant predictor of ToM after controlling for WM and IC.

Table 3. Regression analyses summary for Model 1, including WM and IC as predictors for ToM and Model 2, including WM, IC, nested language, and nested action as predictors for ToM.

	Model 1				Model 2			
Variable	β	*SE*	*t*	*p*	β	*SE*	*t*	*p*
Intercept	0.199	0.271	0.734	.464	0.434	0.354	1.226	.224
WM	2.579	0.670	3.685	<.001	1.361	0.793	1.716	.089

---

[1] Following discussions at a conference we run a separate Pearson correlation between the double nested language task and the hard tree task, as these two tasks might tax more on the ability of NSP due to the two levels of embedding. The correlation was run on the original (non-imputed dataset) and turned out significant (N=117, $r = .195$, $p = 0.036$). When controlled for WM the correlation was not significant anymore (N=115, $r = .027$, $p = 0.782$)

IC	-0.311	0.320	-0.973	.333	-0.416	0.325	-1.278	.206
Nested action					0.282	0.087	3.248	.001
Nested language					0.440	0.467	0.941	.349

Note: adjusted R^2 = .101 (Model 1) and .182 (Model 2)

Exploratory mediation analyses

In order to further assess the relation found between nested action and ToM and to find out whether the relation between age and ToM is mediated by NSP in the action domain an exploratory mediation analyses was run on the imputed datasets. The pooled results showed that the effect of age on ToM was fully mediated via nested action processing (see Figure 3). The indirect effect was tested using percentile bootstrapping with 1.000 samples and turned out significant, suggesting that the older the children the better they were in NSP in the action domain and, thus, the better in the false-belief tasks.

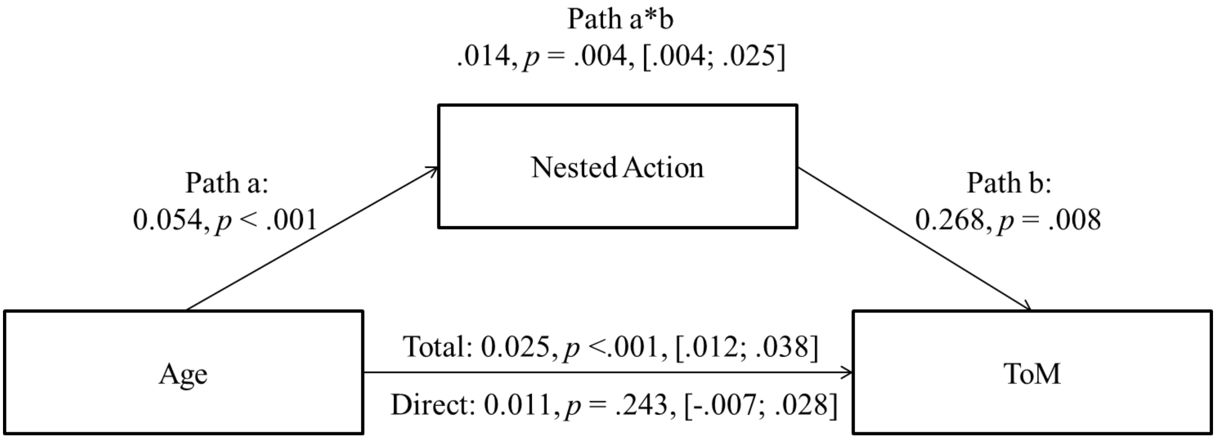


Figure 3. Indirect age effect on ToM through NSP in the action domain with parameter estimates. Numbers in brackets depict 95% confidence intervals.

In a second mediation analysis we controlled for WM to examine whether NSP explains age effects in ToM beyond WM (see Figure 4). Again, the age effects on ToM were fully mediated via nested action processing. While controlling for WM, age predicted NSP in the action domain, which in turn predicted ToM. The indirect effect was tested using percentile bootstrapping with 1.000 samples and turned out significant, suggesting that the older the children, the better they performed in the NSP action task and, thus, the better they were in the false-belief tasks, even when controlling for WM.

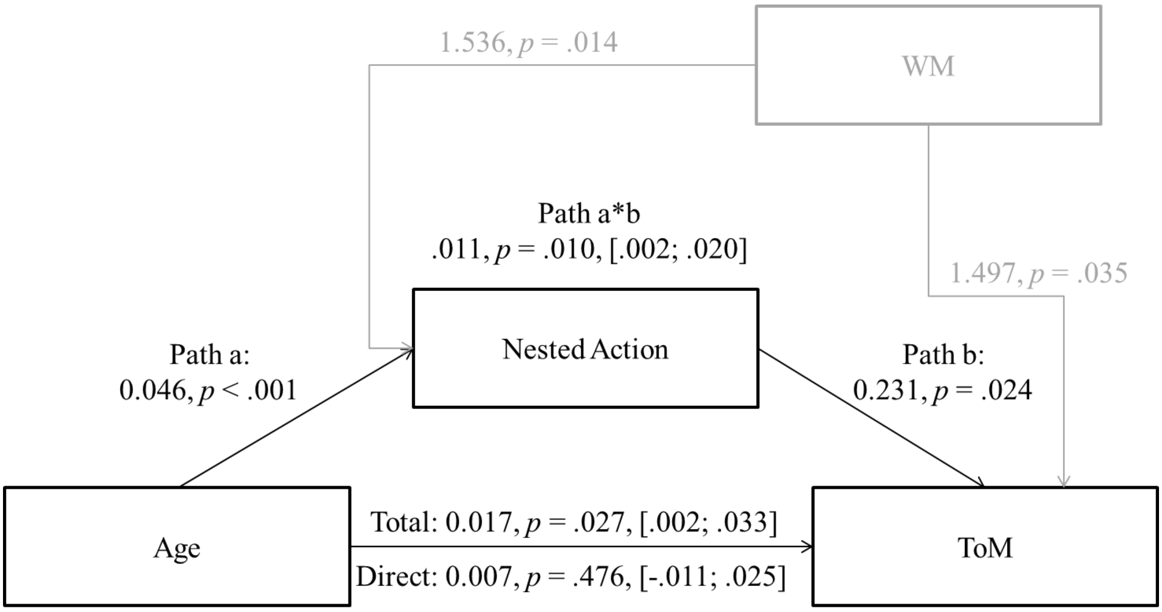


Figure 4. Indirect age effect on ToM through NSP in the action domain with parameter estimates. Numbers in brackets depict 95% confidence intervals. To get the estimates conditional on WM, WM (in grey) was added as covariate to Path a and Path b.

Discussion

Influential theories have proposed that nested structures constitute an important characteristic of human cognition in various domains (e.g., Corballis, 2014; Dawkins, 1976; Lashley, 1951; G. A. Miller et al., 1960). It has been claimed that NSP might constitute a domain-general capacity (Fadiga et al., 2009; Gönül et al., 2018; Grossman, 1980; Jeon, 2014; Lashley, 1951; Marcus, 2006) playing a central role in cognitive development (Corballis, 2014; Greenfield, 1991). In order to empirically examine these claims, the current study investigated whether NSP develops as domain-general capacity and to which extent general cognitive abilities (WM and IC) account for this capacity. Furthermore, we examined claims that NSP plays a role for the early development of ToM (Corballis, 2014; Frye et al., 1995). We assessed three- to six-year old’s performance in various nested structure tasks in the action and language domain, in two general cognitive ability tasks (WM and IC), and in two false-belief tasks. The results showed that NSP in the action and language domain was correlated. The correlation vanished when controlling for WM. These findings are suggestive for a domain-general development of NSP and indicate that WM constitutes an essential basis for it. Moreover, we found NSP to predict ToM and to explain age effects in ToM even when controlling for WM and IC. This is in line with theoretical considerations on the importance of NSP for the early development of ToM (Corballis, 2014; Frye et al., 1995). A number of findings are noteworthy and will be discussed in more detail.

We hypothesized that if NSP develops as domain-general capacity we would find a related development across the action and language domain. Moreover, if WM and IC constitute the basis for processing nested structures but if NSP is a capacity beyond these general cognitive abilities, the relation across the action and language domain should remain but be lowered when controlling for WM/IC. Extending previous work that only focussed on single domains, one central finding was that NSP correlated across the action and language domain. Participants who showed higher NSP scores in the action tasks also performed better in language NSP, indicating similar developmental pathways. This extends previous research that focussed on the domains separately and showed significant improvements during preschool years in NSP within the language (e.g., Corrêa, 1995; Fengler et al., 2016; Kidd & Bavin, 2002; Villiers et al., 1979) and action domain (e.g., Bello et al., 2014; Flynn & Whiten, 2008; Freier et al., 2015, 2017; Greenfield & Schneider, 1977; Melzel & Paulus, 2021) by directly comparing the development of NSP across two different domains.

In a second step, we explored the role of general-cognitive functions for the developing NSP capacity. Interestingly, IC and WM correlated with both, nested language and action. Whereas the correlation between the domains was reduced in strength but remained significant when controlling for IC, it was not significant anymore when controlling for WM. The current study adds to previous findings that explored the relevance of WM and

IC for NSP within the domains (e.g., Fengler et al., 2016; Ferrigno et al., 2020; Meyer et al., 2013) by investigating the role of WM/IC across different domains of NSP. Most importantly, WM—being crucial to maintain hierarchically higher-level information while processing nested lower-level information—seems to explain the parallel developmental pathways across the action and language domain. Taken these results in isolation, they speak against the development of a domain-general NSP capacity beyond WM and might suggest that children's developmental improvement in the processing of nested structures could be conceptualized as a specific form of WM performance.

However, taking the findings of NSP and ToM into account a more complex picture emerges, indicating that it might be too early for such a conclusion: NSP predicted ToM beyond WM and IC and explained the age effects of ToM even when controlling for WM. Hence, the developmental changes in NSP cannot be solely explained by a growth in WM. The complete picture thus also entails some evidence for the development of a domain-general NSP capacity beyond WM. We will further discuss these findings in the following paragraphs.

Our second key question was whether NSP plays a role for the early development of ToM. Most relevant, NSP in the action domain predicted ToM beyond WM and IC and explained the age effects of ToM even when controlling for WM. Our study extends previous work pointing to parallel developments in NSP and ToM (Frye et al., 1995; Grosse Wiesmann, Schreiber, et al., 2017) and supports theoretical proposals that children's ToM development is based on their ability to apply nested rules (Corballis, 2014; Frye et al., 1995). Whereas previous studies focused on the overlap of neural correlates of hierarchical processing and ToM (Grosse Wiesmann, Schreiber, et al., 2017) and showed that WM and IC relate to false-belief understanding (Devine & Hughes, 2014; Grosse Wiesmann, Friederici, et al., 2017; Milligan et al., 2007; Moses, 2005) our study is the first to show that NSP is—beyond WM and IC—important for the early development of ToM.

One possible explanation for the ambiguous findings that WM explained the correlation between language and action whereas NSP explained the age effects of ToM beyond WM might be related to the different strengths of the underlying representations. It should be noted that the language task focused on sentence understanding while in the action domain carrying out or predicting actions was paramount. In line with previous theoretical considerations on graded representations (see Munakata, 2001) producing and applying nested structures (as in the action and ToM tasks) might require and reflect stronger representations than comprehending nested structures (as in the language tasks). In the same vein there is evidence that comprehension precedes production (e.g., Benedict, 1979; Fraser et al., 1963). In the current study, nested structure representations in the language tasks might have been—opposed to the representations in the action domain and in ToM—not strong enough to unravel the domain-general NSP capacity beyond WM. If this holds true, it might also explain the finding that, in contrast to NSP in the action domain, NSP in the language domain was not a significant predictor of ToM when controlling for WM/IC.

One strength of the current study is that, in order to asses NSP, we relied on a number of established tasks (Fengler et al., 2016; Freier et al., 2017; Greenfield & Schneider, 1977; Melzel & Paulus, 2021). To build aggregated scores for NSP in the two domains we relied on an exploratory factor analyses in the action domain and computed the mean over the two tasks in the language domain. Indeed, we found that the tasks in the respective domains were meaningfully related: a factor analysis confirmed the existence of a single factor in the action domain and the two language tasks were significantly correlated. Taken together, by using various tasks we increased the likelihood of assessing NSP rather than any task specifics and the results suggest that the data on which our main analyses were run constitute a reliable basis.

It is important to note that the current study focused on the development of nested opposed to sequential structure processing. Nested structures are characterized by hierarchically lower units that are nested into hierarchically higher elements, so that the hierarchically higher elements span over the lower units. Thus, the processing requires maintaining the information of the higher elements during processing the nested elements. In contrast, elements of a sequence can be processed in a linear order. Previous research showed that pre-schoolers quickly learn to perform and imitate novel sequential actions (e.g., Buchsbaum et al., 2011; Yanaoka & Saito, 2019) whereas the findings of the current study as well as of former studies point to a NSP development during preschool years (Fengler et al., 2016; Flynn & Whiten, 2008; Freier et al., 2015, 2017; Kidd & Bavin, 2002; Villiers et al., 1979). One open question concerns how we can explain the development of NSP. One possibility is to consider the role of internal reflection. More precisely, Allen and colleagues (Allen et al., 2021; Allen & Bickhard, 2018) suggested that in the course of the preschool period children acquire a domain-general cognitive ability to reflect, which in turn enables higher-order cognition. Yet, one could also argue that hierarchical thinking actually supports the development of reflection; or that both abilities are so deeply intertwined that there is no causal primacy of one above the other. Another possibility is to consider the role of conversations and parental scaffolding in the constitution of higher-order cognitive abilities (cf. Carpendale & Lewis, 2004).

Finally, theoretical claims of Karmiloff-Smith (1997) could provide a helpful framework. She proposes that the child initially acquires procedural knowledge which is stored in rather inflexible, implicit, sequential representations and that these representations become more flexible, accessible, and manipulable by the process of representational redescription. Whereas the former representations seem sufficient to process linear sequences, the latter seem pivotal to process the various levels of nested structures while maintaining the hierarchically higher elements. Thus, children's later development of processing nested compared to sequential structures might be traced back to the process of representational redescription.

Our study is the first to point to the development of a domain-general NSP capacity beyond general cognitive abilities and to show that NSP is important for the early development of ToM. Nevertheless, there are some limitations that should be mentioned. First, we did not assess general language abilities. Thus, their impact on the current findings remains unclear. Yet, we solely relied on tasks that have previously been assessed with children in the same age range (Fengler et al., 2016; Freier et al., 2017; Gönül et al., 2018; Greenfield & Schneider, 1977; Melzel & Paulus, 2021). Thus, it seems unlikely that younger participants did not understand the tasks. Second, it remains unclear how the current findings relate to the development of processing nested structures in other domains. Whereas the current study focused on nested actions, sentences, and false-belief understanding nested structures have been claimed to also play a role in the visio-spatial domain (Martins et al., 2014), music (Koelsch et al., 2013; Martins et al., 2017), and tool-use (Byrne & Russon, 1998; Gönül et al., 2018). In future research, it would be interesting to investigate the relation to these other domains as well.

In conclusion, we found similar developmental pathways of nested structure processing across the action and language domain. While this relation was explained by working memory, nested structure processing predicted Theory of Mind beyond working memory and inhibitory control and explained the age effects of Theory of Mind even when controlling for working memory. This might suggest the development of a domain-general capacity of nested structure processing beyond general cognitive abilities and points to its important role for false-belief understanding.

References

Allen, J. W., & Bickhard, M. H. (2018). Stage fright: Internal reflection as a domain general enabling constraint on the emergence of explicit thought. *Cognitive Development*, *45*, 77–91. https://doi.org/10.1016/j.cogdev.2017.12.005

Allen, J. W., Çelik, B., & Bickhard, M. H. (2021). Age 4 transitions: Reflection as a domain-general development for explicit reasoning. *Cognitive Development*, *59*, 101091. https://doi.org/10.1016/j.cogdev.2021.101091

Astington, J. W., & Jenkins, J. M. (1999). A longitudinal study of the relation between language and theory-of-mind development. *Developmental Psychology*, *35*(5), 1311–1320. https://doi.org/10.1037/0012-1649.35.5.1311

Bello, A., Sparaci, L., Stefanini, S., Boria, S., Volterra, V., & Rizzolatti, G. (2014). A developmental study on children's capacity to ascribe goals and intentions to others. *Developmental Psychology*, *50*(2), 504–513. https://doi.org/10.1037/a0033375

Benedict, H. (1979). Early lexical development: Comprehension and production. *Journal of Child Language*, *6*(2), 183–200. https://doi.org/10.1017/s0305000900002245

Botvinick, M. M. (2008). Hierarchical models of behavior and prefrontal function. *Trends in Cognitive Sciences*, *12*(5), 201–208. https://doi.org/10.1016/j.tics.2008.02.009

Buchsbaum, D., Gopnik, A., Griffiths, T. L., & Shafto, P. (2011). Children's imitation of causal action sequences is influenced by statistical and pedagogical evidence. *Cognition*, *120*(3), 331–340. https://doi.org/10.1016/j.cognition.2010.12.001

Byrne, R. W., & Russon, A. E. (1998). Learning by imitation: A hierarchical approach. *Behavioral and Brain Sciences*, *21*(5), 667-684. https://doi.org/10.1017/s0140525x98001745

Carpendale, J. I. M., & Lewis, C. (2004). Constructing an understanding of mind: The development of children's social understanding within social interaction. *Behavioral and Brain Sciences*, *27*(1), 79-96. https://doi.org/10.1017/s0140525x04000032

Chomsky, N. (1956). Three models for the description of language. *IRE Transactions on Information Theory*, *2*(3), 113–124. https://doi.org/10.1109/TIT.1956.1056813

Conway, C. M., & Christiansen, M. H. (2001). Sequential learning in non-human primates. *Trends in Cognitive Sciences*, *5*(12), 539–546. https://doi.org/10.1016/S1364-6613(00)01800-3

Cooper, R. P., Ruh, N., & Mareschal, D. (2014). The goal circuit model: A hierarchical multi-route model of the acquisition and control of routine sequential action in humans. *Cognitive Science*, *38*(2), 244–274. https://doi.org/10.1111/cogs.12067

Corballis, M. C. (2014). *The recursive mind: The origins of human language, thought, and civilization*. Princeton University Press.

Corrêa, L. M. S. (1995). An alternative assessment of children's comprehension of relative clauses. *Journal of Psycholinguistic Research*, *24*(3), 183–203. https://doi.org/10.1007/BF02145355

Dawkins, R. (1976). Hierarchical organisation: A candidate principle for ethology. In P. P. G. Bateson & R. A. Hinde (Eds.), *Growing points in ethology.* Cambridge University Press.

Devine, R. T., & Hughes, C. (2014). Relations between false belief understanding and executive function in early childhood: A meta-analysis. *Child Development*, *85*(5), 1777–1794. https://doi.org/10.1111/cdev.12237

DiStefano, C., Zhu, M., & Mîndrilã, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research, and Evaluation*, *14*(1), Article 20. https://doi.org/10.7275/da8t-4g52

Fadiga, L., Craighero, L., & D'Ausilio, A. (2009). Broca's area in language, action, and music. *Annals of the New York Academy of Sciences*, *1169*, 448–458. https://doi.org/10.1111/j.1749-6632.2009.04582.x

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Fengler, A., Meyer, L., & Friederici, A. D. (2015). Brain structural correlates of complex sentence comprehension in children. *Developmental Cognitive Neuroscience*, *15*, 48–57. https://doi.org/10.1016/j.dcn.2015.09.004

Fengler, A., Meyer, L., & Friederici, A. D. (2016). How the brain attunes to sentence processing: Relating behavior, structure, and function. *NeuroImage*, *129*, 268–278. https://doi.org/10.1016/j.neuroimage.2016.01.012

Ferrigno, S., Cheyette, S. J., Piantadosi, S. T., & Cantlon, J. F. (2020). Recursive sequence generation in monkeys, children, U.S. Adults, and native Amazonians. *Science Advances*, *6*(26), eaaz1002. https://doi.org/10.1126/sciadv.aaz1002

Fitch, W. T., & Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science*, *303*(5656), 377–380. https://doi.org/10.1126/science.1089401

Flavell, J. H. (2004). Theory-of-mind development: Retrospect and prospect. *Merrill-Palmer Quarterly*, *50*(3), 274–290. https://www.jstor.org/stable/23096166

Flynn, E., & Whiten, A. (2008). Imitation of hierarchical structure versus component details of complex actions by 3- and 5-year olds. *Journal of Experimental Child Psychology*, *101*(4), 228–240. https://doi.org/10.1016/.jecp.2008.05.009

Fraser, C., Bellugi, U., & Brown, R. (1963). Control of grammar in imitation, comprehension, and production. *Journal of Verbal Learning and Verbal Behavior*, *2*(2), 121–135. https://doi.org/10.1016/S0022-5371(63)80076-6

Freier, L., Cooper, R. P., & Mareschal, D. (2015). The planning and execution of natural sequential actions in the preschool years. *Cognition*, *144*, 58–66. https://doi.org/10.1016/j.cognition.2015.07.005

Freier, L., Cooper, R. P., & Mareschal, D. (2017). Preschool children's control of action outcomes. *Developmental Science*, *20*(2), e12354. https://doi.org/10.1111/desc.12354

Friederici, A. D., Bahlmann, J., Friedrich, R., & Makuuchi, M. (2011). The neural basis of recursion and complex syntactic hierarchy. *Biolinguistics*, *5*(1-2), 87–104.

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development*, *10*(4), 483–527. https://doi.org/10.1016/0885-2014(95)90024-1

Gönül, G., Takmaz, E. K., Hohenberger, A., & Corballis, M. (2018). The cognitive ontogeny of tool making in children: The role of inhibition and hierarchical structuring. *Journal of Experimental Child Psychology*, *173*, 222–238. https://doi.org/10.1016/j.jecp.2018.03.017

Greenfield, P. M. (1991). Language, tools and brain: The ontogeny and phylogeny of hierarchically organized sequential behavior. *Behavioral and Brain Sciences*, *14*(04), 531–595. https://doi.org/10.1017/S0140525X00071235

Greenfield, P. M., & Schneider, L. (1977). Building a tree structure: The development of hierarchical complexity and interrupted strategies in children's construction activity. *Developmental Psychology*, *13*(4), 299–313. https://doi.org/10.1037/0012-1649.13.4.299

Grosse Wiesmann, C., Friederici, A. D., Singer, T., & Steinbeis, N. (2017). Implicit and explicit false belief development in preschool children. *Developmental Science*, *20*(5), e12445. https://doi.org/10.1111/desc.12445

Grosse Wiesmann, C., Schreiber, J., Singer, T., Steinbeis, N., & Friederici, A. D. (2017). White matter maturation is associated with the emergence of Theory of Mind in early childhood. *Nature Communications*, *8*, 14692. https://doi.org/10.1038/ncomms14692

Grossman, M. (1980). A central processor for hierarchically-structured material: Evidence from Broca's aphasia. *Neuropsychologia*, *18*(3), 299–308. https://doi.org/10.1016/0028-3932(80)90125-6

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, *298*(5598), 1569–1579. https://doi.org/10.1126/science.298.5598.1569

Hofer, A., & Aschersleben, G. (2005). *'Theory of Mind'-Skala für 3- bis 5-jährige Kinder* (2nd ed.). Max Planck Institut für Kognitions- und Neurowissenschaften, Arbeitsbereich Psychologie.

Jeon, H. A. (2014). Hierarchical processing in the prefrontal cortex in a variety of cognitive domains. *Frontiers in Systems Neuroscience*, *8*, 223. https://doi.org/10.3389/fnsys.2014.00223

Kail, R. V., Lervåg, A., & Hulme, C. (2016). Longitudinal evidence linking processing speed to the development of reasoning. *Developmental Science*, *19*(6), 1067–1074. https://doi.org/10.1111/desc.12352

Karmiloff-Smith, A. (1997). *Beyond modularity: A developmental perspective on cognitive science* (3rd printing). *Learning, development, and conceptual change*. MIT Press.

Kaufman, A. S., & Kaufman, N. L. (2015). *KABC-II* ((P. Melchers & M. Melchers, Trans.)) (German Edition). Pearson.

Kazak, A. E. (2018). Editorial: Journal article reporting standards. *The American Psychologist*, *73*(1), 1–2. https://doi.org/10.1037/amp0000263

Kidd, E., & Bavin, E. L. (2002). English-speaking children's comprehension of relative clauses: evidence for general-cognitive and language-specific constraints on development. *Journal of Psycholinguistic Research*, *31*(6), 599–617. https://doi.org/10.1023/A:1021265021141

Koelsch, S., Rohrmeier, M., Torrecuso, R., & Jentschke, S. (2013). Processing of hierarchical syntactic structure in music. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(38), 15443–15448. https://doi.org/10.1073/pnas.1300272110

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jefress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–136). New York: Wiley.

Lindner, K. (2003). The development of sentence-interpretation strategies in monolingual German-learning children with and without specific language impairment. *Linguistics*, *41*(2), 213–254. https://doi.org/10.1515/ling.2003.008

Maffongelli, L., D'Ausilio, A., Fadiga, L., & Daum, M. M. (2019). The ontogenesis of action syntax. *Collabra: Psychology*, *5*(1), 21. https://doi.org/10.1525/collabra.215

Marcus, G. F. (2006). Cognitive architecture and descent with modification. *Cognition*, *101*(2), 443–465. https://doi.org/10.1016/j.cognition.2006.04.009

Martins, M. J. D., Gingras, B., Puig-Waldmueller, E., & Fitch, W. T. (2017). Cognitive representation of "musical fractals": Processing hierarchy and recursion in the auditory domain. *Cognition*, *161*, 31–45. https://doi.org/10.1016/j.cognition.2017.01.001

Martins, M. J. D., Laaha, S., Freiberger, E. M., Choi, S., & Fitch, W. T. (2014). How children perceive fractals: Hierarchical self-similarity and cognitive development. *Cognition*, *133*(1), 10–24. https://doi.org/10.1016/j.cognition.2014.05.010

Melzel, S., & Paulus, M. (2021). The development of the prediction of complex actions in early childhood. *European Journal of Developmental Psychology*, *18*(2), 161–183. https://doi.org/10.1080/17405629.2020.1773786

Meyer, L., Obleser, J., & Friederici, A. D. (2013). Left parietal alpha enhancement during working memory-intensive sentence processing. *Cortex*, *49*(3), 711–721. https://doi.org/10.1016/j.cortex.2012.03.006

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior*. Henry Holt and Co. https://doi.org/10.1037/10039-000

Miller, S. A. (2009). Children's understanding of second-order mental states. *Psychological Bulletin*, *135*(5), 749–773. https://doi.org/10.1037/a0016854

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.

Moses, L. J. (2005). Executive functioning and children’s theories of mind. In B. F. Malle & S. D. Hodges (Eds.), *Other minds: How humans bridge the divide between self and others.* Guilford Press.

Munakata, Y. (2001). Graded representations in behavioral dissociations. *Trends in Cognitive Sciences*, *5*(7), 309–315. https://doi.org/10.1016/S1364-6613(00)01682-X

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(4), 515–526.

Pulvermüller, F., & Fadiga, L. (2010). Active perception: Sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, *11*(5), 351–360. https://doi.org/10.1038/nrn2811

Richland, L. E., & Burchinal, M. R. (2013). Early executive function predicts reasoning development. *Psychological Science*, *24*(1), 87–92. https://doi.org/10.1177/0956797612450883

Ryle, G., & Tanney, J. (2009). *The Concept of Mind: 60th Anniversary Edition*. Routledge.

Simon, H. A. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society*, *106*(6), 467–482. https://www.jstor.org/stable/985254

Simpson, A., & Riggs, K. J [Kevin J.] (2005). Inhibitory and working memory demands of the day-night task in children. *British Journal of Developmental Psychology*, *23*(3), 471–486. https://doi.org/10.1348/026151005X28712

Villiers, J. G. de, Flusberg, H. B. T., Hakuta, K., & Cohen, M. (1979). Children's comprehension of relative clauses. *Journal of Psycholinguistic Research*, *8*(5), 499–518. https://doi.org/10.1007/BF01067332

Villiers, J. G. de, & Villiers, P. A. de. (2000). Linguistic determinism and the understanding of false beliefs. In P. Mitchell & K. J. Riggs (Eds.), *Children's reasoning and the mind* (pp. 191–226). Psychology Press.

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

Yanaoka, K., & Saito, S. (2019). Repeated sequential action by young children: Developmental changes in representational flexibility of task context. *Developmental Psychology*, *55*(4), 780–792. https://doi.org/10.1037/dev0000678

Zheng, X., Swanson, H. L., & Marcoulides, G. A. (2011). Working memory components as predictors of children's mathematical word problem solving. *Journal of Experimental Child Psychology*, *110*(4), 481–498. https://doi.org/10.1016/j.jecp.2011.06.001