Deeds not Words?
Dynamics in Word and Action Learning during Early Childhood

Sarah Eiteljörge

Play is the highest form of research.
- Albert Einstein

Deeds not Words?
Dynamics in Word and Action Learning during Early Childhood

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades
"Doctor rerum naturalium"

der Georg-August-Universität Göttingen
im Promotionsprogramm Behavior and Cognition
der Georg-August University School of Science (GAUSS)

vorgelegt von
Sarah Fé Vivian Eiteljörge
aus Berlin

Göttingen, Juli 2019

Betreuungsausschuss

Prof. Dr. Nivedita Mani
Psychologie der Sprache
Georg-August-Universität Göttingen

Prof. Dr. Birgit Elsner
Entwicklungspsychologie
Universität Potsdam

Prof. Dr. Alexander Gail
Sonsorimotor Group
Georg-August-Universität Göttingen

Prüfungskommission

Prof. Dr. Nivedita Mani
Psychologie der Sprache
Georg-August-Universität Göttingen

Prof. Dr. Birgit Elsner
Entwicklungspsychologie
Universität Potsdam

Weitere Mitglieder der Prüfungskommission

Prof. Dr. Alexander Gail
Sonsorimotor Group
Georg-August-Universität Göttingen

Dr. Tanya Behne
Entwicklungspsychologie
Georg-August-Universität Göttingen

Prof. Dr. Markus Steinbach
Germanistische Linguistik
Georg-August-Universität Göttingen

Prof. Dr. Annekathrin Schacht
Affektive Neurowissenschaft und Psychophysiologie
Georg-August-Universität Göttingen

Tag der mündlichen Prüfung: Freitag, der 20.09.2019

Deeds not Words?
Dynamics in Word and Action Learning during Early Childhood

Dissertation

to acquire the doctoral degree in mathematics and natural science
"Doctor rerum naturalium"

at the Georg-August-Universität Göttingen
within the doctoral degree programme Behavior and Cognition
of the Georg-August University School of Science (GAUSS)

submitted by
Sarah Fé Vivian Eiteljörge
from Berlin, Germany

Göttingen, July 2019

Acknowledgments

First and foremost, I want to thank my supervisors Nivi and Birgit, as well as my travelling companions Vivien and Maurits, for their enduring intellectual and motivational support during this exciting journey.

Nivi, you have been an inspirational role model over the past years and I admire your curiosity, your constant willingness to learn, and your humanity in leadership.

Further, I want to thank all the families for visiting us in the WortSchatzInsel, my colleagues for the engaging discussions, and all the amazing student research assistants for helping out with everything. Without all of you, it would not have been possible to test so many participants in the last three to four years and to look at development across such a wide age range. I also want to thank Alexander Gail for his input during the thesis committees which broadened my perspective and let me see my work in the bigger picture of cognition. Additionally, I want to to thank Crossing the borders (FOR 2253), the RTG 2070 "Understanding Social Relationships", and the Leibniz-ScienceCampus "Primate Cognition" for financial support and great workshops, and all their members for interesting exchanges and entertainment. And I want to thank the Sprawi for all the support during the years and for providing my favourite workplace.

I also want to thank my family. You let me explore my interests and supported me whenever I needed it.

Last but not least, I want to thank you, Sebastian, for discussing statistics with me late at night, for asking annoying questions which always brought me further along the way, for calming me down by just being there, and for (swing) dancing with me through the seasons.

Abstract

Children grow up in a complex, multimodal environment and need to extract information from this environment to learn about the world. When caregivers interact with their children, they often use words and actions concurrently, and this might influence children's processing of words and actions. In this thesis, we investigated children's and adults' learning of words and actions for objects in multimodal contexts to understand the interactions between the child and the environment during learning in early development.

In the first study, we explored the developmental trajectory of word and action learning. 12-month-olds learned to associate actions but not words with objects, 24-month-olds learned neither word- nor action-object associations, 36-month-olds learned to associate words but not actions with objects, and adults learned to associate both words and actions with objects. Thus, children developed from learning action-object to word-object to both associations. A possible reason for this observation could be that the participants' individual preferences influenced their word and action learning differently across age groups. In the second study, we therefore investigated how the learner's own preferences can influence learning of words and actions for objects. We replicated the findings from Study 1 showing that 24- to 36-month-olds learned word-object but not action-object associations, while adults learned both associations. Furthermore, children's preferences influenced their target looking in the word-object, but not in the action-object condition, and adults' preferences influenced target looking in the action, but not in the word condition. Thus, participants' preferences influenced their processing of information but did not influence learning. In addition to the individual preferences, how the two domains are presented might influence learning. Thus, in the third study, we investigated the cross-domain influences between words and actions during children's word learning. Between 30 and 42 months, consistency in the co-occurrence of actions in contrast to variability in the co-occurrence supported children's word-

object association learning. Adults learned such associations independent of the actions presented. Accordingly, the structure of the input influenced learning in toddlerhood but its influence decreased with age.

In summary, this thesis sheds light on the trajectory of word and action learning in development, how the learner's preferences influence learning, and how words and actions influence each other during learning. Therefore, this thesis illustrates interactions between the learner, her changing interests and competences, and the environment, and its form and complexity, and thereby reflects the dynamic development of learning. As such, this thesis provides supporting evidence for theories that postulate interactions between different levels of an organism and the environment across time as the basis of development.

Zusammenfassung

Kinder wachsen in einer komplexen, multimodalen Umgebung auf und müssen Informationen aus dieser Umgebung extrahieren, um Wissen über die Welt zu erlangen. Wenn Bezugspersonen mit ihren Kindern interagieren, verknüpfen sie oft Wörter und Handlungen mit Objekten. Dies kann Auswirkungen darauf haben, wie Kinder die Wörter und Handlungen verarbeiten. In dieser Arbeit untersuchten wir das Lernen von Kindern in multimodalen Kontexten, in denen sowohl Wörter als auch Handlungen für Objekte präsentiert wurden, um Interaktionen zwischen dem Kind und der Umgebung während des Lernens in der frühen Entwicklung zu verstehen.

In der ersten Studie untersuchten wir die Entwicklung des Wort- und Handlungslernens. Zwölfmonatige lernten, Handlungen, aber keine Wörter, mit Objekten zu assoziieren, 24-monatige lernten weder Wort- noch Handlung-Objekt-Assoziationen, 36-monatige lernten, Wörter, aber keine Handlungen mit Objekten zu assoziieren, und Erwachsene lernten, Wörter und Handlungen mit Objekten zu assoziieren. Kinder scheinen also zunächst Handlung-Objekt- und dann Wort-Objekt-Assoziationen zu lernen, bevor sie beides lernen. Dabei bleibt jedoch unberücksichtigt, dass die individuellen Präferenzen der Probanden das Lernen von Wörtern und Handlungen beeinflussen könnten. In der zweiten Studie untersuchten wir daher, wie die Präferenzen des Lernenden das Erlernen von Wörtern und Handlungen für Objekte beeinflussen können. Wir replizierten die Ergebnisse der ersten Studie, die zeigen, dass 24- bis 36-monatige Kinder Wort-Objekt-, aber keine Handlung-Objekt-Assoziationen lernten, während Erwachsene beide Assoziationen lernten. Die Präferenz der Kinder beeinflusste ihr target looking bei Wort-Objekt- aber nicht bei Handlung-Objekt-Assoziationen, und die Präferenz der Erwachsenen beeinflusste deren target looking bei Handlung-Objekt- aber nicht bei Wort-Objekt-Assoziationen. Die Präferenzen schienen also das Verarbeiten von Informationen zu beeinflussen, nicht aber das Lernen. Zusätzlich zu den individuellen Präferenzen

kann die Darstellungsweise der beiden Domänen das Lernen beeinflussen. In der dritten Studie untersuchten wir daher die domänenübergreifenden Einflüsse zwischen Wörtern und Handlungen beim Erlernen von Wort-Objekt-Assoziationen. Bei Kindern im Alter zwischen 30 und 42 Monaten unterstützte es das Lernen von Wort-Objekt-Assoziationen, wenn die gleichzeitig auftredenden Handlungen konsistent und nicht variabel waren. Erwachsene lernten solche Assoziationen unabhängig von den begleitenden Handlungen. Die Struktur des Inputs hatte also einen Einfluss auf das Lernen im Kleinkindalter, aber dieser Einfluss schwindet mit dem Alter.

Zusammenfassend beleuchtet diese Dissertation den Verlauf des Wort- und Handlungslernens in der Kindesentwicklung, wie die Präferenzen des Lernenden diesen Verlauf beeinflussen, und wie sich Wörter und Handlungen während des Lernens gegenseitig beeinflussen. In dieser Arbeit werden daher die Interaktionen zwischen der Lernenden, ihren sich ändernden Interessen und Kompetenzen, und der Form und Komplexität ihrer Umgebung dargestellt, was zusammen die dynamische Entwicklung des Lernens reflektiert. Als solche liefert diese Arbeit Belege für Theorien, die Wechselwirkungen zwischen verschiedenen Ebenen eines Organismus und der Umgebung im Verlaufe der Zeit als die Basis für Entwicklung postulieren.

Contents

Introduction **1**
Communication . . . 3
Action . . . 5
Parallels and differences in word and action learning . . . 7
Domain-general vs. domain-specific mechanisms . . . 12
Dynamic approaches to development . . . 14
Consistency and variability . . . 15
Theories on word and action learning . . . 17
How does word-object and action-object association learning develop? . . . 21
Do preferences influence learning of word-object and action-object associations? . . . 22
Are there cross-domain influences between the word and the action domain during learning? . . . 24
How can we evaluate learning? . . . 29
How can we measure preferences? . . . 30
Analyses of eye-tracking data . . . 32
Outline . . . 35

Word-object and action-object association learning across early development **36**
Abstract . . . 36
Introduction . . . 37
Methods . . . 42
Participants . . . 42
Stimuli . . . 43
Training phase stimuli . . . 43
Test phase stimuli . . . 44
Procedure . . . 44
Experimental design . . . 45
Pre-processing . . . 47
Data analysis . . . 49
Results . . . 51
Discussion . . . 56

Do preferences for words and actions influence word-object and action-object learning in early childhood? **64**
Abstract . . . 64
Introduction . . . 65
Current study . . . 69
Methods . . . 71
Participants . . . 71
Material . . . 71
Procedure . . . 73
Experimental design . . . 74
Pre-processing . . . 76

Results 77
Discussion 88

Consistency of co-occurring actions influences young children's word learning **94**
Abstract 94
Introduction 95
Methods 101
Participants 101
Stimuli 101
Procedure 103
Experimental design 103
Preprocessing 105
Results 106
Discussion 113

Discussion **121**
Parallels and differences in word and action learning 123
The influence of learners' preferences 124
Consistency and variability 127
Methods and statistics 130
Theories on word and action learning 131
Domain-general vs. domain-specific mechanisms 137
Conclusion 141

Other Contributions **143**
Unimodal word and action learning at 12 and 24 months 143
The impact of verbal cues on 14-and 24-month-olds' visual action anticipation 145

References **147**

Declaration **173**

Appendix **174**
Study 1: Word-object and action-object association learning across early development 174
Study 2: Do preferences for words and actions influence word-object and action-object learning in early childhood? 178
Study 3: Consistency of co-occurring actions influences young children's word learning 181

List of Tables

1 Intersensory Redundancy Hypothesis 19
2 Study 1: Learning and test phase 47
3 Study 1: Descriptives of baseline-corrected target looking 51
4 Study 2: Descriptives of action violation phase 85
5 Study 2: Descriptives of side violation phase 86
6 Study 3: Descriptives of baseline-corrected target looking 107
7 Study 1: GLMM testing differences between conditions 174
8 Study 1: GLMM testing differences within conditions 175
9 Study 2: LMM testing differences between conditions and age groups 178
10 Study 2: LMM testing the interaction of preference and target looking 178
11 Study 2: GLMM testing differences between conditions and age groups 179
12 Study 3: GLMM testing differences between conditions 181
13 Study 3: GLMM testing differences within conditions 182

List of Figures

1 Streamer plot 34
2 Study 1: Stimuli 43
3 Study 1: Time course of children's target looking 55
4 Study 1: Time course of adults' target looking 56
5 Study 2: Stimuli 72
6 Study 2: Time course of conditions for both age groups 80
7 Study 2: Time course of children's target looking 82
8 Study 2: Time course of adults' target looking 84
9 Study 2: Correlation of preference and exposure phase 88
10 Study 3: Stimuli 102
11 Study 3: Training phase 104
12 Study 3: Test phase 105
13 Study 3: Reprint of streamer plot 108
14 Study 3: Time course of participants' target looking 114
15 Study 4: Time course of proportional target looking 144

Introduction

Every human being grows up in a complex world and is required to extract information from the environment to learn and thrive in our society. One important aspect in this context is the communication with others to transfer information between individuals and groups. Another important aspect is the manipulation of objects as can be observed in tool use or play. Caregivers often use both words and actions for objects concurrently in their interactions with children, and this co-occurrence might influence children's processing of information from the two domains.

Children's play in early childhood involves learning both words and actions for objects. For example, children need to learn the word *bunny* and the action *hopping* for a rabbit toy. Children learn words at a rapid pace with their lexicon encompassing around 300 words around their second birthday (Frank, Braginsky, Yurovsky, & Marchman, 2017). Further, they learn to explore objects with their hands, to imitate others' actions, and to use objects to accomplish goals at a similarly young age (e.g., Jones, 2009, Rochat, 1989). Importantly, these two domains are often encountered together: Research has shown that caregivers often use words and actions for objects together when playing with their child (Gogate, Bahrick, & Watson, 2000), and that this combination of information can help the child to learn (Gogate, Maganti, & Laing, 2013). Furthermore, both words and actions share similar structures and prerequisites, require learning, and remain relevant throughout the lifetime.

As such, the comparison of children's word and action learning allows us to investigate similarities and differences in the processing of two domains that share certain structures, co-occur in the environment, and play an important role in children's knowledge acquisition and development. In previous studies, these parallels have motivated researchers to investigate underlying mechanisms in learning of the two domains, and thereby address the question whether such underlying mechanisms appear to be domain-specific or domain-general which would inform theories on

the cognitive structure of human beings (e.g., Childers & Tomasello, 2002, Dysart, Mather, & Riggs, 2016). However, from these studies, it remains unclear how word and action learning develop in early childhood, and uncovering this trajectory is crucial for understanding the domain-generality or -specificity of such mechanisms.

Furthermore, recent theories of development assume multidirectional interactions between a learner and the environment (Lickliter & Honeycutt, 2003, Smith & Thelen, 2003). Evidence for or against such accounts can only be provided by research that specifically addresses interactions between the learner and the environment. In this regard, the development of word and action learning provides an ideal case to study such interactions because they are two domains, that might interact with each other, are co-present in the child's environment, and relevant for the child to learn. Given that children's processing abilities are still developing, how do children accomplish to learn in such multimodal environments? How do children and their environment interact with each other? Which type of information do children extract from their surrounding? And which factors, both on the side of the child and on the side of the environment, can influence this behaviour?

The work in this dissertation sheds light on how children's word and action learning develops, how individual preferences of the learner can influence learning in these domains, and in turn, how the structure of the multimodal input can shape the child's learning behaviour. Thus, this research can help us understand how children integrate information from both the word and the action domain into their current knowledge, and how these two domains interact with the learner in early development. Furthermore, this thesis addresses the question on domain-general and domain-specific mechanisms underlying learning throughout early childhood, and how the learner and the environment interact during learning. Together, the insights from this research enrich our understanding of how children cope with different domains in the multimodal environment early in development and how they become skilled learners over the first few years.

Communication

Communication is an important aspect of evolution, has been described as a tool to share information between members of a group, and can be found in different human and nonhuman species (Tomasello, 2008). The transfer of knowledge is usually accomplished through the transmission of a signal that is sent from the signaller to the receiver. This signal can take on different shapes (e.g., verbal or nonverbal), can be transmitted in different forms (e.g., vocal production, writing), and requires certain behaviours from both the signaller and the receiver: while the signaller has to plan and produce the signal, the receiver has to process and interpret the signal.

To be able to communicate, members of a group are required to learn the specific communication system. For humans, most aspects of language learning can typically be observed in the first decade of life. An integral part to language learning is the subcomponent of word learning because words allow us to refer to concepts and express relations. Word learning describes the acquisition of lexical information with words representing the smallest meaningful unit in language. Accordingly, the acquisition of words allows us to build up a repertoire of units that we can use in our communication with others. These words can either be functional (e.g., when they express relations) or referential (e.g., when they refer to an entity in the world). In the case of referential words, the learner is required to learn the association between the word and the concept (e.g., an object). Since we rely on the integration of new concepts into our current knowledge state throughout life, also the ability to learn words for such concepts remains relevant over the course of a lifetime.

During their first year of life, children develop many skills they require to be amazing word learners. Already immediately after birth, infants show a preference for the voice of the mother (DeCasper & Fifer, 1980), for familiar texts (DeCasper & Spence, 1986), and for their native language (Mehler et al., 1988), suggesting that already in the womb, they learn about the melody and stress patterns of their

native language. Then, over the first ten months of life, infants learn to discriminate between sounds that belong or do not belong to their native language (Werker & Tees, 1984) which helps them to focus on the speech sounds that are relevant to their native phonetic inventory. Further, they learn to segment the continuous stream of speech into words using for example probabilities between words (Saffran, Aslin, & Newport, 1996), prosodic features (Mattys, Jusczyk, Luce, & Morgan, 1999), and repetition (Ota & Skarabela, 2018). Around six months, they can already segment their own name and highly familiar words like *mummy* and *daddy* from the continuous speech stream (Bortfeld, Morgan, Golinkoff, & Rathbun, 2005, Mandel, Jusczyk, & Pisoni, 1995), and this ability will be extended to other words with their growing vocabulary size (Jusczyk & Aslin, 1995). Around the same age, infants start to show image recognition for words such as *mommy* and *daddy*, body parts, and food items (Bergelson & Swingley, 2012, Tincoff & Jusczyk, 1999, 2011).

Later on during their second year of life, their already existing knowledge can help learning. For example, studies have shown that phonologically similar words help target recognition around 18 month (Mani & Plunkett, 2010) but hinder at 24 months due to children's increased lexicon size (Mani & Plunkett, 2011), whereas semantically similar words help target recognition around this age, reflecting structure and restructuring in children's early lexicon (Arias-Trejo & Plunkett, 2010, 2013). In addition, certain biases help children learn new words: for example, 18-month-olds can learn novel word-object associations by excluding already known objects as potential referents (Bloom, 2002, Markman & Wachtel, 1988, Markman, Wasow, & Hansen, 2003). Further, if objects share for example colour (Mani, Johnson, McQueen, & Huettig, 2013), shape (Samuelson, 2002, Smith, 2004), or the same context (Horst, 2013), this knowledge can help them to identify the target object of a new word. By 30 months, children can understand and produce on average around 550 words with a range of 250 to 650 (Frank et al., 2017), which shows that they are well-equipped word learners by that age.

Action

Another important aspect of evolution is the physical manipulation of objects by an agent as reflected in actions and tool use, and which is shared among human and nonhuman species (Tomasello, 1999). Actions can take on different forms and functions: For example, they can be arbitrary actions which can imply a repetitive movement but do not involve a defined end point. They can also be goal-directed actions which are only exhibited to achieve a certain predefined end point. Such actions are especially relevant for tool use since they often relate to the function of an object and serve to accomplish a goal. Furthermore, actions can be described as small units that, together, can form one longer action sequence. Any type of action in this regard then requires an agent who performs the action, and an object on which the action is performed.

To be able to process and perform actions, the individual has to learn the accordingly required skills, and for humans, many aspects of action learning can typically be observed in the first decade of life. The acquisition of actions allows us to build up a repertoire of abilities that we can use for example in object exploration, play, and tool use. Like words, actions can be functional (e.g., when they help to accomplish something) or referential (e.g., when they refer to an entity in the world). In both cases, the learner is required to learn the association between the action and the concept (e.g., an object). And like the ability to learn words, learning new actions, for example, to explore a new object or to use a new tool, remains relevant throughout the lifetime.

At a very young age, children develop into skilled action learners. Already in the womb, a child might feel the movement of the mother and moves by itself (Verbruggen et al., 2018). Between two and five months of age, infants' own object manipulation develops from an oral to a more visual and tactile exploration (Rochat, 1989). Also around five months, infants are able to discriminate and remember actions such as brushing teeth (Bahrick, Gogate, & Ruiz, 2002), although it is still

under discussion to which extend infants can discriminate between matching and mismatching action-object associations (e.g., a woman using a hairbrush to brush her hair vs. her teeth at 5.5 months Bahrick et al., 2002, and six months, Hunnius & Bekkering, 2010). Around six months (but not at four months), infants can visually anticipate the goal of a grasping action (Kanakogi & Itakura, 2011), and direct their attention to object manipulations guided by characteristics in caregivers' play behaviour (Koterba & Iverson, 2009, Lansink & Richards, 1997, in contrast to chimpanzees, Bard & Vauclair, 1984). Around 12 months, children use statistical regularities (Paulus et al., 2011), goal salience (Henrichs, Elsner, Elsner, & Gredebäck, 2012), and behavioural cues (Adam, Reitenbach, & Elsner, 2017, Biro, 2013) to process actions.

Studies with older children have often focused on children's imitation behaviour to understand their action interpretations (see Jones, 2009, for an extensive review on children's imitation and its relevance for dynamic systems theories). For example, from 14 months, children imitate rational actions more than irrational actions (Carpenter, Akhtar, & Tomasello, 1998, Gergely, Bekkering, & Király, 2001), suggesting that children from that age have a rudimentary understanding of others' intentions.

Thus, both in the word and in the action learning literature, researchers have often investigated children's learning in these domains. This has brought about fascinating insights into children's early development in learning words and actions, and shows how quickly they become expert learners in the respective field. These studies focus on specific questions with regard to the respective domain, and thereby, cannot (and do not try to) answer how words and actions interact with each other during learning. When comparing studies in both domains, it might be possible to see influences of certain mechanisms in both domains (e.g., attention). But to study whether actions and words share a similar trajectory in development, maybe share similar mechanisms, or potentially influence each other during learning, we need to

consider studies that directly compare and contrast word and action learning in the same learning environment.

Parallels and differences in word and action learning

From an evolutionary perspective, both domains are important to the development of the modern human primate, and share certain requirements and characteristics across species. From a developmental perspective, both domains can be observed in caregivers' behaviour towards infants and in children's own play and learning. The two domains share hierarchical structures of individual sequences that, together, can form a meaningful unity. To process and express information, it is required in both of these domains to plan, to store information, to keep a goal in mind, to solve problems, to perform actions, among many other cognitive skills. In contrast to communication, actions do not necessarily possess a communicative intent, but due to their structure have been considered a potential precursor to communication.

Human and nonhuman animals alike possess certain characteristics and show certain behaviours that have been coined as precursors to communication. These include for example memory, categorisation (in pigeons, Vauclair, 1996), transitive inferences (see Vasconcelos, 2008, for a review), numerosity (Dehaene, Dehaene-Lambertz, & Cohen, 1998), sequence learning (Terrace, Son, & Brannon, 2003), and social competence (complex social groups in some nonhuman primate species, D. Cheney, Seyfarth, & Smuts, 1986). Different kinds of communication have been described for nonhuman primates (e.g., warning calls in vervet monkeys, D. L. Cheney & Seyfarth, 1981), in birds (e.g., canary song to attract mates, Gardner, Naef, & Nottebohm, 2005), and honeybees (dance to indicate food locations, Von Frisch, 1967) among others. To different degrees, these communication systems involve information about acoustic properties (phonetics), the function of such properties (phonology), structure and function of individual units (morphology), the meaning

of such units (semantics), the relation between small and larger units (syntax), and the bigger context in which communication is placed (pragmatics).

As for communication, human and nonhuman animals alike are required to possess certain characteristics and abilities for actions such as memory, planning, and sequence learning (see Badets & Osiurak, 2015, for example). As described above, they can be found in many different species. Further, actions and tool use have been observed in human and nonhuman primates (Tomasello, 1999, Torigoe, 1985), birds (e.g., New Caledonian crows, Hunt, 1996, Kenward, Rutz, Weir, & Kacelnik, 2006), and mammals (e.g., dolphins, Mann et al., 2008). In nonhuman species, mostly tool use has been investigated and has been found to facilitate for example food retrieval (e.g., Bentley-Condit & Smith, 2010, Seed & Byrne, 2010, Van Schaik, Deaner, & Merrill, 1999). For humans, many different forms of object manipulation can be observed: for example, they use exploratory play to learn about the form and function of objects (e.g., Cook, Goodman, & Schulz, 2011), they use arbitrary action such as *shaking* to guide another person's attention (Matatyaho & Gogate, 2008), and engage in elaborate tool use (Johnson-Frey, 2004).

One large system of actions that involves communication is the field of referential gestures. Gestures can be divided into deictic gestures, such as *show*, *offer*, or *point*, and representational gestures, such as culturally evolved arbitrary signals like *thumbs up*, or iconic gestures such as pantomime and object feature depiction like forming the hand to represent a rabbit (Liszkowski, 2008). Of these different types, only deictic gestures like *show* and *offer* typically include objects. Liszkowski (2008) noted that these gestures involve aspects that can be interpreted as communicative signals such as guiding the attention of a partner, but they could also have evolved from object-oriented, exploratory actions that are not considered communicative (but potentially considered communicative by the interaction partner). Thus, a large part of the literature on gestures focusses on deictic and representational

gestures that are used with communicative intent and that do not necessarily involve objects.

In this thesis, I will predominantly refer to the literature on arbitrary actions, and more specifically, to the literature on action-object association learning because the actions employed in the following studies involve agent-guided, arbitrary object manipulation, and focus on participants' learning of these action-object associations.

The two domains have been studied extensively using observational and experimental studies with children. Studies have shown that both words and actions are often simultaneously present and temporally aligned in the child's environment (Gogate et al., 2000, Suanda, Smith, & Yu, 2016). This synchrony diminishes over the second and third year of life but is nonetheless present even at 30 months (76 % at five to nine months, 59 % at nine to 17 months, 36 % at 21 to 30 months, Gogate et al., 2000). These studies mirror the close connection between words and action in children's environment, and reflect how omnipresent their co-occurrence is, especially during the first year. Studies with older children often find that there are few differences between word and action learning (Childers & Tomasello, 2002, 2003, Dysart et al., 2016, Riggs, Mather, Hyde, & Simpson, 2015). For example, Childers and Tomasello (2002) presented groups of 30-month-olds with either novel noun-object associations ("Look at this! This is a wuggy. See? It's a wuggy.") or novel action-object associations ("Look at this! Look what we can do with this. See? Look what we can do with it."). After the training phase, they tested children's comprehension in a referent selection task and tested children's production by asking them to name an object or to act out an action of an object. They found that children's comprehension for words and actions did not differ significantly across groups, suggesting that children learned the association. Further, Childers and Tomasello (2003) showed that 30-month-olds extended novel actions to similar objects as they did with novel nouns. Later, at 3 to 4 years, Riggs et al. (2015) found that children chose a novel object compared to a familiar object upon hear-

ing a novel word or seeing a novel action, that they retained both associations, and extended both novel words and novel actions to other category members. Further, Dysart et al. (2016) showed that 3- to 4-year-olds' word and action learning was guided by attention to novelty. Together, these studies show similarities between word and action learning in children's development from 30 months on.

In contrast, children's production of novel words and actions for objects seems to be more differentiated. Childers and Tomasello (2002) found in their study that 30-month-old children reproduced actions more often than nouns. A similar study by Hahn and Gershkoff-Stowe (2010) found that 2- and 3-year-olds showed better learning of actions compared to words, better word comprehension compared to word production, and better action production than action comprehension (at two years) or comparable action comprehension and production (at three years). In a second study, they presented 4- to 5-year-olds with both a word and an action for an object, and replicated their previous results. Regarding the differences between comprehension and production, Childers and Tomasello (2002) argue that this might relate to how we perceive novel functions of objects, which are then directly stored as a feature of that object. However, other researchers have similarly argued that words are features of objects (Gliozzi, Mayor, Hu, & Plunkett, 2009). Hahn and Gershkoff-Stowe (2010) claim that the physical interaction with the object might lead to a representation that is easier to access in memory, and therefore leads to better recall. However, this only explains discrepancies between word and action production, but not discrepancies between comprehension and production of actions they find at two years. Hence, this area of research requires more investigation in the future.

Thus, except for production, children's word and action learning show a remarkably similar development above 30 months. However, in most of these studies, children learned either a word or an action for the object, and were not presented with both types of information simultaneously but rather in an unimodal context

(although language was present in the action conditions). This approach contrasts with the studies on caregivers' behaviour showing that the word and the action for the same object were often presented simultaneously. Accordingly, a bi- or multimodal study design would mirror caregivers' behaviour more closely, which was implemented in the second study with 4- to 5-year-olds by Hahn and Gershkoff-Stowe (2010) described above. But how does children's learning develop before the preschool age?

While research on the development of word and action learning in multimodal contexts is rare, hypotheses can be drawn from studies investigating multisensory information. In these studies, researchers have looked at participants' processing of auditory and visual stimuli in uni- and bimodal settings (Barnhart, Rivera, & Robinson, 2018, Robinson & Sloutsky, 2004, Sloutsky & Napolitano, 2003). For example, Robinson and Sloutsky (2004) found that infants between eight and 16 months showed an auditory compared to visual preference, 4-year-olds fluctuated between both, and adults showed a visual preference while processing both types of information. Furthermore, Sloutsky and Napolitano (2003) showed that 4-year-olds fluctuated only when both modalities were presented simultaneously. If presented individually, a higher percentage of 4-year-olds processed auditory but not visual information. Barnhart et al. (2018) replicated these findings but showed a more nuanced picture with regard to younger (around 20 years) and older (around 80 years) adults: They found that younger adults showed patterns of auditory dominance whereas older adults showed a pattern of visual dominance. When applied to word and action learning, these results suggest that in early development children might attend to auditory information, i.e., language, more readily than to visual information, i.e., actions.

As described above, most studies have looked at word and action domains separately, although the similarities and differences between the two domains could inform us about underlying mechanisms both evolutionarily and developmentally

as I will argue below. If the two domains share their underlying mechanisms as suggested by studies on both words and actions (Childers & Tomasello, 2002, 2003, Dysart et al., 2016), such mechanisms might also drive learning in both of these domains.

Domain-general vs. domain-specific mechanisms

As reviewed above, both the word and the action domain share a common structure, common requirements, and depend on learning. Is this learning in early development based on general mechanisms that are shared across domains or are such mechanisms relevant to specific domains and not to others? It has been an on-going debate whether certain cognitive abilities require domain-general or domain-specific processes, and answers to this question could inform theories on the cognitive structure of humans.

In 1965, Chomsky argued that children's language learning is accomplished at such a high rate despite the poor input that it seems impossible that children acquire language at some point in life. Thus, each human being must have a specific knowledge set that provides some basic linguistic mechanisms or information, which can then be adapted to the native language once the child is presented with language-specific information (e.g., as suggested in "Government and Binding", and "Minimalism", Chomsky, 1993, 2015). This framework of modules has been adapted to other cognitively demanding tasks such as representing others' mental states (Baron-Cohen, 1997), resulting in the idea that human beings are born with many different modules that are domain-specific, and thus, only relevant to one specific cognitive ability (e.g., Fodor, 1983). This nativist approach requires that these modules are hard-wired in some way, for example, in the genes. However, this would require an evolutionary process in which all of these modules developed as gene mutations (e.g., Lickliter & Honeycutt, 2003). Such processes require an extensive amount of time which seems unrealistic with respect to the comparatively fast

evolutionary development of, e.g., linguistic abilities in humans (see Baronchelli, Chater, Pastor-Satorras, & Christiansen, 2012, for a theoretical approach).

In contrast, other researchers have emphasised the importance of the environment (Piaget, 1952, see Elman, 2005, for an overview of connectionist models of development). Their view is that children possess only very basic innate structures, such as domain-general learning mechanisms. These then help to acquire cognitive abilities, such as language, which emerge through the child's general predispositions and her interaction with the environment. In this view, complex multidirectional interactions between the organism (i.e., the learner) and the environment account for development. Tomasello (1999) has argued from an evolutionary perspective that the development of language (and similarly advanced cognitive skills) are not based on hard-wired processes, but that this development rather requires several small changes that result in a larger synergy effect, which then allows for these abilities. For Tomasello (1999) among many others, one such aspect is the development of cumulative culture which brought about joint attention, imitative learning, and linguistic symbols and cognitive representations, and thereby induced nonlinear change in human development.

Karmiloff-Smith (1992) suggested a synthesis of domain-specific and domain-general approaches to development: in her view, the modularity of the mind is shaped through the child's experience. While the child possesses some form of biological predispositions for learning and some implicit representations (e.g., numerosity, rule following), only through the interaction with the environment can these representations become explicit (e.g., counting, metalinguistic awareness). In such a framework, processes in the brain become more domain-specific with time and experience (Karmiloff-Smith, 1992).

There are several hypotheses that we can derive from the ideas presented above. On the one hand, if two domains (i.e., words and actions) develop in synchrony, we could assume that a) they require similar mechanisms, b) they require different

mechanisms which develop at the same time, or c) they require different mechanisms but have a shared mediator that makes the two domains appear as if they developed in synchrony (i.e., categorisation of information, selective attention). On the other hand, if two domains do not develop in synchrony, we could assume that a) they require different mechanisms or b) they require similar mechanisms but have a shared mediator that weighs information differently, and thereby, guides individual development. This question can help us understand how human beings learn, whether learning requires certain mechanisms that are shared across domains (e.g., Elman, 2005), or whether children possess (e.g., Chomsky, 1993) or develop (Karmiloff-Smith, 1992) specific learning mechanisms in the two domains.

Dynamic approaches to development

With the emergence of new research methods and the motivation to understand both evolutionary and developmental change, researchers have recently aspired to combine research from embryology, molecular genetics, phylogeny, and evolutionary biology into one research area called evolutionary developmental biology (in short, evo-devo, Gilbert, Opitz, & Raff, 1996, Müller, 2007). This broad combination of areas allows researchers to study development on different time scales, from microbiology to evolution, and study the relations between these scales. More specifically, this approach attempts to model complex dynamic interactions between the organism and its environment, and how this interaction causes change in the organism's system. Lickliter and Honeycutt (2003) described how this new perspective on evolution and development has direct consequences for theories in research areas such as neuroscience, cognitive psychology, and linguistics. They wrote that the integration of evo-devo in current theories on cognition is inevitable for biologically plausible explanations of development. The authors conclude by writing that their framework transcends previously assumed dichotomies, and thereby acknowledges development as an interactive process on many levels. Accordingly, their dynamic

approach combines an evolutionary and a contemporary view on development, including a phylogenetic and an ontogenetic perspective.

When we consider a dynamic approach to development, such as presented in the work by Lickliter and Honeycutt (2003), which predetermines complex interactions between the organism (i.e., the learner) and the environment, we imply that different aspects not only could develop in parallel and require similar mechanisms, but that these different aspects also influence each other. These influences can be found on many levels within the organism, within the environment, from organism to environment, and from environment to organism (Lickliter & Honeycutt, 2003).

With respect to words and actions, both domains can be found in the complex, multimodal environment of the learner. Accordingly, interactions in the environment and with the learner can result in synergy effects of these different domains that can either facilitate learning (i.e., supportive structures), be indifferent to learning (i.e., neutral structures), or hinder learning (competitive structures). How these influences unfold, then depends on the state of the learner and the form of the input.

Consistency and variability

Information entropy is one scale according to which the form of the input can differ. Information entropy describes the average density of information in a given unit: Low entropy describes a state of more repetition and therefore less information content, whereas high entropy describes a state of less repetition and therefore more information content (Shannon, 1948).

Consistency and variability can be described as part of this scale that begins with pure repetition of the same information (zero entropy), continues with increasing levels of differences, and ends in a state where all possibilities are equally likely (highest entropy). In reference to this scale and in the context of our studies, pure repetition can be exemplified as the presentation of the *same* stimuli, consistency can be described as the presentation of the *same kind* of stimuli, and variability can

be described as the presentation of *different kinds* of stimuli with varying degree, which is maximal in the chaos scenario. This description of entropy implies that there are different statistical regularities at play for consistency and variability. A by-product of consistency is a statistically high co-occurrence of information: if a word, an action, and an object always appear together, their statistical co-occurrence is high but it might not be clear whether the word refers to the action or the object because both associations are equally probable. This ambiguity can only be resolved when, for example, the sentence structure defines the function of the word (e.g., noun vs. verb). In contrast, variability implies a statistically low co-occurrence of events, and therefore can help to disentangle information: if a word and an object always appear together, but the accompanying action is always different, the statistical probability that the word refers to the object is higher than the probability that the word refers to the action. Note that both consistency and variability are not one specific point on the spectrum, but can cover different ranges of entropy, and therefore also different degrees of statistical co-occurrence.

With regard to learning, consistency and variability have been considered to serve different functions. Consistency is often assumed to deepen and consolidate knowledge due to the higher number in repetition. In contrast, variability is assumed to broaden and diversify knowledge due to the higher information content. Further, variability can lead to better abstraction and generalisation of the acquired knowledge (Perry, Samuelson, Malloy, & Schiffer, 2010). Exactly because of these two different functions, both consistency and variability are important to learning, and in combination, can lead to stable and diverse knowledge (Goldenberg & Sandhofer, 2013, Vlach & Sandhofer, 2011). However, entropy is solely a concept specific to the structure of the input but how do children learn from the input?

Theories on word and action learning

Common theories of development on how children learn words and actions often describe these two processes separately: Such theories are based on one single domain (e.g., word learning), and incorporate different areas within this domain (e.g., phonetic, phonological, syntactic, and semantic information). For example, theories on language acquisition have tried to model speech perception (Jusczyc, 1997) or the development from speech perception to word learning (Werker & Curtin, 2005). Similarly, theories on action development have often tried to model children's action learning: For example, researchers have modelled children's processing of (goal-directed) actions with regard to their interpretation of partners as intentional agents (e.g., Gergely & Csibra, 2003), rationality of the action, and action anticipation (Gredebäck & Melinder, 2010).

Due to their focus, these theories can be highly specific, explicit, and elaborate regarding the details of the model, and thus, are well formulated for hypothesis generation. However, they often lack the possibility to include information from other domains, and can therefore neither predict nor model any relationships between them. In the context of learning from multimodal input, and in the light of the ideas of evo-devo, it seems ecologically valid to turn to theories that try to incorporate information from different senses and domains, and accordingly, might be able to model any similarities, differences and cross-domain influences in word and action learning.

The Intersensory Redundancy Hypothesis is one such framework in which information from different senses and domains is combined into one system. Gibson (1966) proposed that the different senses form the perceptual system. The combined senses then allow for the detection of commonalities in the input across the senses. More explicitly, Bahrick, Lickliter, and Flom (2004) wrote that two types of information overlap can be described: amodal stimulus information is redundant information perceived with more than one sense (i.e., clapping provides redundant

information such as tempo in the acoustic and the visual domain) whereas modality-specific information is only perceived by one of the senses (e.g., the hands' movement during clapping can only be perceived visually). They provide evidence that both nonhuman and human infants detect amodal relations, for example, with regard to temporal and spatial information (see Lickliter & Bahrick, 2000, for an overview). In the input, this information can either be presented unimodally (i.e., clapping in the distance can only be perceived auditorily) or it can be presented multimodally (i.e., hands' movement can be perceived visually and acoustic effect can be perceived auditorily).

Based on the idea that redundant information perceived by the different senses can help the learner process and integrate information, Bahrick and Lickliter (2002) suggested that the combination of stimulus property (amodal vs. modality-specific) and input property (uni- vs. multimodal) then forms a 2×2 matrix of possible hypotheses (see Tab. 1): amodal information in a multimodal context and modality-specific information in a unimodal context should facilitate processing due to the redundant information provided in these scenarios. In contrast, amodal information in a unimodal context and modality-specific information in a multimodal context should attenuate processing. This difference should vanish during development with the increasing perceptual abilities of the child. Overall, intersensory redundancy could, thus, be taken as a means to organise information, and thereby, help to guide attention, perception, and cognition across species and development (Bahrick et al., 2004).

Ecological and dynamic theories of learning have taken a similar stance regarding the integration of complex information but focus on the aspect of learning, and the interaction between the learner and the environment, rather than perception. Already in 1914, Montessori described the child as an active, exploring individual who interacts with the environment: in her view, the child seeks learning environments in which the abilities and interests of the child and the information in the

Table 1
Hypotheses based on the Intersensory Redundancy Hypothesis. Table adapted from Bahrick and Lickliter (2002).

		Stimulus property	
		Amodal	Modality-specific
Input property	Multimodal (auditory-visual)	facilitated	attenuated
	Unimodal (auditory or visual)	attenuated	facilitated

input are optimally aligned (coined as the "Goldilocks effect" in the more recent literature, see for example, Kidd, Piantadosi, & Aslin, 2012). This harmonisation of learner and environment then allows for contented learners and optimal learning outcomes (and hopefully confident and contented members of society, Montessori, 1914).

In more recent formulations of these ideas, Smith and Thelen (2003) argued in their overview that children's impressively fast development is not necessarily due to innate mental structures, but precisely emerges through the dynamic interactions between the child and the environment (Karmiloff-Smith, 1992, Thelen & Smith, 1996). In their view, all aspects in a learning situation can potentially interact with each other and no specific aspect is dominant per se in the learning situation. Accordingly, interactions between the learner (and the learner's abilities and interests) and the environment (and its form and complexity) shape the learning experiences (Hirsh-Pasek, Golinkoff, & Hollich, 2000, Smith & Thelen, 1993).

Further, Smith and Thelen (2003) write that development can then be seen as a series of events differing in stability: transitions in learning are marked by instability in the system, but are intertwined with robust, stable phases in which the learned aspect is consolidated. Accordingly, variability in the learning setting should lead to instability in the system, which in turn leads to a cascade of so-called phase shifts (Stephen & Dixon, 2008). These phase shifts result in qualitative and quantitative changes in the system, and according to dynamic systems approaches,

such changes reflect learning (Karmiloff-Smith, 1992, Piaget, 1952, Thelen & Smith, 1996, Twomey, Ma, & Westermann, 2017). Importantly, these processes are not established in a certain, pre-defined sequence in time, but happen on many different levels and timescales in the organism, and across development and evolution. Accordingly, this dynamic approach to learning situates itself in the dynamic framework of evolutionary developmental biology, with a focus on cognitive development.

While the general framework has been formulated in rather broad terms, adaptations to specific areas can show more concrete implementations of the basic ideas. For example, Gogate and Hollich (2013) adapted the framework to children's milestones in word learning. They suggest that these milestones can be described as sensitive periods of increased interaction between the learner and the environment, above described as *phase shifts*. If timings in these dynamic events change, this affects the developmental trajectory of the learner, as is often the case with atypically developing children (e.g., children with cochlear implants or born preterm, Gogate & Hollich, 2013). Accordingly, the interplay between the environment and the learner across time shapes the learning experience.

With regard to word and action learning, such a dynamic systems theory is rather broad (in contrast to single-domain theories), and therefore allows to incorporate children's processing and learning of different domains in one theoretical framework. Due to its dynamic nature, this framework allows to ground aspects of learning in the developing perceptual abilities of the learner, to acknowledge interactions within the input, and between the input and the learner. More specifically, it makes direct predictions regarding consistency and variability in the input on the one hand and learning outcomes on the other hand: consistency would be associated with the stable phases between shifts, and variability would be associated with unstable phases during shifts. In summary, while this framework appears to be broad, and therefore less specific concerning individual research questions, it allows for a

more inclusive perspective on child development with regard to word and action learning.

How does word-object and action-object association learning develop?

As described above, studies investigating children's word and action learning suggest that children's learning behaviour across the two domains appears to be similar, with the authors of such studies concluding that this supports the idea of shared underlying mechanisms (Childers & Tomasello, 2002, 2003, Dysart et al., 2016). These studies have focused on children older than 30 months and often presented them with either the word domain or with the action domain (but see Hahn & Gershkoff-Stowe, 2010). In a context that provides both types of information (Gogate et al., 2000), processing of both domains might be more challenging at younger ages (Robinson & Sloutsky, 2004), and thus, children might focus on either words or actions.

Based on this research, we investigated whether children and adults learn words or actions for objects when presented with both domains simultaneously (albeit temporal synchrony), and how this learning develops over early childhood between 12 and 36 months. If word and action learning follow a similar trajectory as described in Robinson and Sloutsky (2004), younger children should learn words (i.e., auditory information) for objects but not actions (i.e., visual information) for objects. This would also be in accordance with other studies which suggest that language predominates in early development (Fulkerson & Waxman, 2007). In later development, around 30 months and beyond, this influence might decrease resulting in similar learning outcomes for words and actions as shown in the studies with older children (Hahn & Gershkoff-Stowe, 2010). This research can help us understand how children become skilled learners in multimodal learning environments and how they learn to integrate different domains into their knowledge.

Further, this research can shed light on the discussion regarding domain-general and domain–specific mechanisms: if word and action learning develop in synchrony, we could assume that first, these domains require similar mechanisms and these mechanisms might therefore be domain-general. Second, they require different mechanisms which develop at the same time scale. Third, they require different mechanisms but share a common mediator. If word and action learning do not develop in synchrony, we could assume that first, they require different mechanisms, or second, they require similar mechanisms but learner-specific characteristics like individual preferences guide development. Note that if word and action learning require different mechanisms, these could then be either domain-general or domain-specific, or a combination of both types of mechanisms.

In this study, we investigated word and action learning at the group level. By including different age groups, we could indirectly evaluate the influence of the developing perceptual abilities of the participants. However, we could not evaluate any differences between the individual children within an age group. For example, it could be that children show differential interest for the two domains. That is, some of the children might be more interested in learning words for objects while others might be more interested in learning actions for objects. This would then lead to an interaction between learning outcomes and the word and action domains, showing a dynamic interaction between the child's personal interests and the learning environment.

Do preferences influence learning of word-object and action-object associations?

Already in the early 20th century, researchers suggested that children's own curiosity and preferences influence their processing and learning of new information (see Kidd & Hayden, 2015, for a review on curiosity and its neurological underpinnings). One central argument has been that children have limited processing

abilities but are confronted with a complex environment (called sampling problem), and thus, any mechanism that reduces the focus of attention to a limited amount of information in the input can help the child to zoom in on environmentally or personally relevant information (Kidd & Hayden, 2015).

Kidd and Hayden (2015) wrote that if we consider a broad definition of curiosity that includes any orientation to new information, even infants show certain behaviours that adhere to this definition, such as orienting to novel stimuli in preference paradigms (Houston-Price & Nakai, 2004), motion as a cue to animacy (Aslin & Shea, 1990), and social information (Baldwin, 1993). Furthermore, infants show systematic patterns of information seeking: infants around seven to eight months showed higher interest in stimuli of medium complexity suggesting that they orient towards input which has the right amount of complexity or novelty given the learner's state of knowledge (Kidd et al., 2012, Kidd, Piantadosi, & Aslin, 2014). This behaviour is thought to result in optimal learning given that it fills the gap between the learner's current knowledge state and understanding (Loewenstein, 1994). This type of information seeking has been modelled computationally revealing similar results (Twomey & Westermann, 2017), and even supports word learning in 3- to 5-year-olds (Nussenbaum & Amso, 2016). Further, this behaviour allows them to reduce uncertainty and to understand causal relationships in their environment (Bonawitz, van Schijndel, Friel, & Schulz, 2012).

Based on this research, we investigated whether children's word and action learning was influenced by their own individual preferences for either words or actions for objects. We hypothesised, based on the idea that curiosity boosts learning, that children who show a preference in one domain would also show better learning in the respective domain in contrast to the other domain. Thus, if a child is more interested in words for objects, we expected this child to also learn a word for an object better than an action. Importantly, this behaviour should only be observable as long as the child requires mechanisms to adapt the information in the input to

his or her processing abilities. Thus, with increasing abilities, preferences should still emerge but not influence learning any longer.

This research helps us understand how characteristics of the child, namely individual preferences, are evident in the word and action domain, and how these preferences contribute to learning in the two domains in early development. Thus, this research question directly addresses the interaction between the provided input and the learner's characteristics. Further, if word and action learning do not develop in synchrony in the first study, we could assume that either they require different mechanisms or they require similar mechanisms but the learner's preferences and other characteristics influence the learning behaviour.

Are there cross-domain influences between the word and the action domain during learning?

As described above, according to dynamic approaches to learning, the learner and the immediate environment interact with each other in the learning situation. Thus, on the one hand, the child brings his or her own characteristics to the learning situation, such as perceptual abilities, memory, or preferences. On the other hand, the input has its own characteristics that it brings into the learning situation, such as the amount, the complexity, or the consistency and variability of the provided information.

Research on cross-domain influences has brought to light a large array of candidates that influences language and action processing. For example, objects can help 9-month-olds to discriminate non-native phonetic categories (Yeung & Werker, 2009), different faces can help sound discrimination even in adults (Mani & Schneider, 2013), and reliability of a partner influences both word learning (Scofield & Behrend, 2008) and action imitation (Zmyj, Buttelmann, Carpenter, & Daum, 2010). Further, cross-domain influences between the language and the action domain have been investigated. For example, caregivers often use physical manipula-

tions during play while naming an object, thereby helping the child to disambiguate the word referent in a complex and rich environment (Gogate et al., 2013, Masur, 1997, Messer, 1978). Further, specific actions like *shaking* seem to facilitate learning of word-object associations for 8-month-old infants (Matatyaho-Bullaro, Gogate, Mason, Cadavid, & Abdel-Mottaleb, 2014), and such action cues still appear to be beneficial for word learning at 14 months (Werker, Cohen, Lloyd, Casasola, & Stager, 1998). Vice versa, language cues can improve children's discrimination of actions (Brand & Tapscott, 2007), children's imitation performance already from 12 months on (Herbert, 2011, Seehagen & Herbert, 2010), and can guide 14-month-olds (Sciutti, Lohan, Gredebäck, Koch, & Rohlfing, 2016), and 3- to 5-year-olds to focus on certain aspects of an action, for example the movement or the goal, and to imitate the respective part (Elsner & Pfeifer, 2012, Pfeifer & Elsner, 2013). However, the influence of verbal information on action processing is not always clear-cut (Adam, Eiteljoerge, Mani, & Elsner, 2018), and can sometimes even hinder such processing (Zack, Gerhardstein, Meltzoff, & Barr, 2013), possibly due to the cognitive demands that are posed on the learner in such a learning context (see also Gampe & Daum, 2014, for interfering influences of language on action perception at 12 months and facilitating influences at 24 months). Nonetheless, further studies have suggested that also motor and language development are not necessarily independent of each other: changes from crawling to walking appear to be related to a change in caregivers' responsive behaviour (Walle, 2016), children's perceptive and productive vocabulary (He, Walle, & Campos, 2015, Walle & Campos, 2014), and their object exploration (Karasik, Tamis-LeMonda, & Adolph, 2011). Thus, it seems that, on the one hand, intersensory information helps processing in both the word and the action domain, potentially guiding children's attention to certain aspects of a situation, and on the other hand, the child's abilities provoke a change in their environment, and thereby influence their own learning in the two domains.

Indeed, studies have found that consistency and variability influence learning differently. Consistency is often assumed to deepen and consolidate knowledge while variability is assumed to broaden and diversify knowledge. For example, consistency in word-object pairings is a prerequisite for learning a word-object association (Smith & Yu, 2008). Consistency in spatial information (Benitez & Smith, 2012, Samuelson, Smith, Perry, & Spencer, 2011), and temporal and linguistic contexts (Roy, Frank, DeCamp, Miller, & Roy, 2015) help word learning. Further, studies on story book reading show that reading always the same book to the child supported novel word learning, in contrast to reading different books with the same novel words (Horst, Parsons, & Bryan, 2011, Williams & Horst, 2014). Also, consistency in the linguistic structure of a sentence supports 2-year-olds in learning new words (Schwab & Lew-Williams, 2016). Together, these studies indicate that consistency in the input can help children acquire new words and consolidate their word knowledge.

However, as described above, consistency implies a statistically high co-occurrence of the same information. When children are presented with objects and actions while they hear a novel word, the referent of this novel word might not be clear from the beginning. It could refer either to the object or to the action. To learn the correct association, this ambiguity has to be resolved using either syntactic information or other contextual cues.

As further described above, variability is considered to broaden knowledge. For example, variability in visual appearance facilitates learning of perceptual categorisation (Mather & Plunkett, 2011, Quinn & Bhatt, 2010), and variability in speakers facilitates learning of similar sounding words (Rost & McMurray, 2009, 2010). Perry et al. (2010) found that 18-month-olds show better abstraction of category knowledge and extension to novel exemplars when trained on highly variable sets of objects, but not when trained on similar sets of objects. Accordingly, the authors

claimed that this broadening is the result of abstraction and generalisation which is induced through the variability in the input.

One study showing that variability can also be helpful in word learning was conducted by Twomey et al. (2017). They found that 2-year-old children only learned word-object associations when the background colour differed across trials compared to a stable white background. The authors suggest that, in terms of dynamic systems accounts, this low level of entropy in the environment was enough to induce instability in the system, and thereby, helped children to learn. Their claim was supported by Twomey, Ranson, and Horst (2014) who showed that 30-month-old children learned category labels when objects differed in colour, but not when the objects differed in colour and shape. Accordingly, the contrast in results might be due to the salience of the variability in the input: a change in background colour might be a rather subtle influence that helps to keep the attention on screen. However, if the content which changes reaches a certain level of salience, it might guide the attention away from the relevant aspect, and thereby exacerbate learning (see also Rost & McMurray, 2009).

Based on this research, we investigated whether the consistency and variability in actions as part of the multimodal input influences children's word learning, and how this develops over early childhood. Since we investigated word learning and not category learning, consolidating instead of generalising knowledge might be required in this learning context. Therefore, children should show better word learning when the actions are consistent across encounters. Furthermore, variability in actions might be rather salient and therefore distract children from learning the words (as in Twomey et al., 2014). However, if the consistent mapping of information across modalities remains ambiguous for the child (for example, if the child cannot yet discriminate between nouns and verbs), variable actions might be more helpful in learning the word-object association as variability leads to a decrease in statistical co-occurrence. Furthermore, if variability also in our context helps to

induce unstable states in the learner in terms of a dynamics system approach, variability should lead to better word learning in early development. Only when children have developed the perceptual abilities to cope with the complex input at hand, might they be able to learn words in both contexts. If however, children's learning is influenced by neither variability nor consistency in the input, their word learning might be generally supported by the presence of actions, since in terms of intersensory redundancy, the additional information about the object might help the child to form a strong representation of the association. Thus, from the literature to-date it is difficult to predict whether children learn words for objects better when actions differ or remain the same across encounters during early development. Only for older children, the evidence converges and suggests that older children should be able to learn words for objects independent of the actions presented.

This research will help us understand how children's word learning is influenced by the structure of the input and by different domains present in a rich learning environment. From a developmental perspective, these interactions between the input and the learner might change with the increasing knowledge and abilities of the learner. Thus, this research question addresses the interaction between the provided input and the learner across development by focusing on the form of the input that is provided.

In summary, the present thesis consists of three studies that investigate learning across different domains in early development: In the first study, we investigated whether children between 12 and 36 months and adults choose to associate words or actions with objects when presented with both. In the second study, we investigated whether the individual preferences of children between 24 and 36 months and adults influence their learning behaviour in the two domains. In the third study, we investigated how the consistency and variability in co-occurring actions can influence learning of a word-object association in children between 18 and 42 months and adults. Together, these studies provide new insights into the commonalities

and differences in learning across different domains, how the learner's interests influences such learning, and how information in the learning context interacts with the learner's behaviour across development.

How can we evaluate learning?

For all three research questions, it is required to evaluate participants' learning behaviour when presented with novel words, actions, and objects. To accomplish this, typical learning studies require a training phase and a test phase. For example, participants go through a training phase in which they encounter novel or unfamiliar objects and are presented with their novel or unfamiliar names. After the training phase, participants go through a test phase in which two or more of the objects are presented on screen, and it is measured whether participants look to the correct object upon hearing its label and not to the distractor object (c.f., Twomey et al., 2017). This measure then reflects participants' proportional target looking in relation to participants' distractor looking (target/(target + distractor)). If, on average, participants look more to the target than to the distractor object, it is assumed that they successfully formed an association between the label and the object. Note however that these word learning effects are usually of short-lived nature given the length of the experiment (often three to ten minutes) and the absent transfer of knowledge to life outside the lab. Thus, although I write of *learning* in this thesis, the underlying process might be a formation of an *association* that is stored temporarily but not in long-term memory, also coined *fast mapping* in the word learning literature (see Horst & Samuelson, 2008, for an example of limited retention abilities in such tasks at 24 months).

To record a person's gaze as a reaction to stimuli, eyetrackers can be mounted below the stimulus display (for example, either underneath a screen or a live play theatre). The eyetracker usually contains a built-in infrared light that illuminates the participant's eyes and an infrared camera that records its reflection. The con-

trast in reflection of light from the cornea is then used to estimate where the participant is looking on the screen. This method allows to study many mechanisms like information processing and learning in tightly controlled experiments. It is especially interesting for studying young children, because typically developing infants' gaze improves drastically in the first three months of age (Banks & Salapatek, 1978). Thus, their gaze can give us insight into what they find interesting and how they learn, long before they have the linguistic and metalinguistic abilities to share this information with us verbally. This also makes the method interesting for other nonverbal populations as long as gaze is normally developed.

For our setup, we used two screens and a remote desktop eyetracker that are connected to the same computer. One screen, outside the experimental booth, was used as the experimenter screen to control the experiment. The second screen inside the booth was used to display the stimuli. A remote desktop eyetracker was placed under the screen in the booth, and was configured with regard to the size of the screen and the distance of the participant to the screen. For the present studies, we used a Tobii X120 eyetracker (Manual, 2008) that records gaze data every 16 ms. For stimuli presentation, we used the software E-Prime and its Tobii extension (Schneider, Eschman, & Zuccolotto, 2002).

How can we measure preferences?

In addition to measuring learning, we needed a measure of participants' preferences to evaluate whether preferences influenced learning. Studies looking at preferences in preverbal infants have often used looking times as a measure (e.g., Cooper & Aslin, 1990). In these studies, a checkerboard is presented on screen while the stimuli are auditorily presented. Here, looks away from the screen indicate that the child lost interest in the stimulus and the next can be played. Longer looking times are then usually taken to reflect higher interest (and less boredom) in a type of stimulus. For example, 1-month-olds showed longer looking times for infant-

directed speech compared to adult-directed speech (Cooper & Aslin, 1990). These studies work well with infants since they usually find the flicker board interesting enough to stay on task. However, this paradigm would probably be difficult with older children since they might be unchallenged in this task. Thus, in the present work, we wanted to provoke an active preference-based choice of the participant while the participant would be engaged in the task.

To accomplish such an interactive design, we developed a gaze-contingent eye-tracking paradigm that allows for a choice between two options. This paradigm requires a programme that evaluates live information of the participant's looking behaviour and will then respond to the participant's gaze accordingly. Thus, in addition to recording gaze data, the data is simultaneously analysed providing a live estimate of where the participant is looking on screen. This information can then be used to present new information on-screen. In such experiments, participants can actively engage in the task.

This paradigm has been used mainly in studies of reading where the appearance of text on-screen changes based on where the participant is looking (see Rayner, 2014, for a review). In more recent years, this paradigm has been adapted to children. For example, Wang et al. (2012) used this paradigm to investigate 6- to 8-month-olds' discovery of agency (i.e., that certain events are contingent on their own actions). They presented a red dot on screen, which, when fixated, would trigger the display of an animal picture on the other side of the screen. The results indicate that even at 6 months of age, infants learned to control the experiment within three trials and showed anticipatory looks to the side of the image display. Furthermore, such designs have been used to demonstrate 8-month-olds' gaze following in the absence of a human being as long as the demonstration is contingent (Deligianni, Senju, Gergely, & Csibra, 2011), provide evidence that 10-month-olds expect certain outcomes of their own actions (Kenward, 2010), and can provide tools to improve attentional control in 11-month-olds (Wass, Porayska-Pomsta, & Johnson, 2011).

Recently, gaze-contingent paradigms have been extended to children diagnosed with autism spectrum disorder to study their reactions to social signals in comparison to typically developing children (Vernetti, Senju, Charman, Johnson, & Gliga, 2018, Vernetti, Smith, & Senju, 2017) and to develop new tools for diagnosis (B. Sullivan, Wilson, & Saldaña, 2019).

Analyses of eye-tracking data

In the studies presented in this thesis, we tested whether participants showed recognition of the correct target object upon hearing its label (or seeing its action) during the test phase, which would indicate that participants associated the word (or the action) with the correct object during the learning phase. To evaluate their target recognition, we measured their proportional target looking during the test phase. Any significant increases compared to chance level would indicate that they successfully formed the association. Most statistical tests that have been used to address this question can be classified into three approaches which come with their own advantages and disadvantages.

In many previous word learning studies that required eyetracking, participants' proportional target looking was collapsed across time and then fed into an analysis of variance (ANOVA) in which the groups means and their variance from different learning conditions were compared. To test learning within one condition, a t-test comparing averaged proportional target looking against chance was usually computed. If significant, participants' target looking was either significantly above (target looking) or significantly below (distractor looking) chance level. If no significant differences arose, it could not be assumed that participants associated the word with the object. This analysis allows for conclusions about participants' learning. However, this type of analysis reduces the resolution of the data due to the loss of temporal information and cannot account for other sources of variance (e.g., item information such as object colour or exact age of the child).

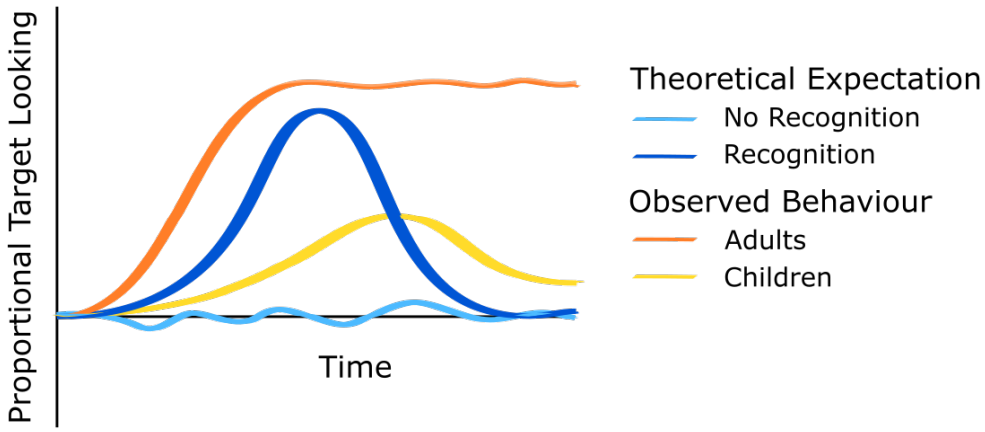
To circumvent the loss in information, Monte Carlo simulations can be used to find significant divergences between conditions in the time course (c.f., Maris & Oostenveld, 2007, Von Holzen & Mani, 2012). In this analysis, significant differences between conditions for each time point are grouped by adjacency, and together, form so-called time clusters. The sum statistic of each cluster is then compared to the sum statistic of a distribution of sum statistics that was obtained through repeated random shuffling of the data. Although this type of analysis incorporates *time*, it can only reflect significant divergences within the trial and therefore does not model target looking throughout the trial.

Another possibility to circumvent the loss of information is to use a generalised linear mixed model (GLMM) that includes *time* as a factor. In comparison to ANOVAs, such models also allow to include more sources of variance. In addition, including *time* then allows to evaluate any effects related to the time course. Especially in educational psychology, these types of analyses, also called growth curve analysis (GCA), have been used to model changes in e.g., academic performance over several years (Campbell, Pungello, Miller-Johnson, Burchinal, & Ramey, 2001, Gutman, Sameroff, & Cole, 2003). However, change over time is often not linear, neither for academic performance nor for target looking. To model different time courses, it is possible to project the time course on orthogonal polynomials, that are based on the factor *time*. The coefficients can be used as factors in a GLMM to model changes in target looking across the time course.

For target looking, we can formulate the following theoretical expectations about participants' proportional target looking (see Fig. 1). If at the group level, participants do not recognise the target object upon hearing its label, their target looking should stay around chance level (light blue). This would then reflect that roughly 50 % of the participants look to the target and 50 % look to the distractor object at any given time point. If at the group level, participants successfully recognise the target object upon hearing its label, their target looking should be at chance

level in the beginning, then increase as they look to the correct object, and then decrease towards chance level again (dark blue). This curve resembles a quadratic function, or a quartic function if tails at the beginning and end of the time course are included. In contrast to these theoretical expectations, adults' target looking usually increases shortly after the start of the time course and then stays high, suggesting that adults fixate the target until the end of the trial if they learned the association in the training phase (orange). Children's target looking often starts slightly later and increases at a weaker and slower pace (yellow). In addition, their target looking is often more variable and therefore requires flexibility in modelling and interpretation. To model these different possibilities, we then require *time* as a linear, quadratic, cubic, and potentially quartic factor.

Figure 1. Streamer plot of theoretical expectations and typical behavioural observations of looking behavior across time. Light and dark blue represent theoretical expectations, yellow and orange represent typical behaviour.



Thus, GLMMs including time and its polynomials take the full time course of the data into account, and therefore, provide high sensitivity and temporal resolution. Further, they can account for additional sources of variance in the data. However, they come at the cost of increased complexity in both mathematical and interpretational terms. Nonetheless, due to the increased sensitivity and resolution, these models allow to evaluate target looking that indicates learning but that might be more subtle and would therefore be obscured by any analysis that requires

non-temporal data. This is especially important for data from young children, in comparison to adults, where learning effects might be weaker and where target looking might not always follow theoretically expected curves.

In this thesis, we used the traditional ANOVA approach to compare our results to previous word learning results in the literature. In addition, we used GLMMs including the factor *time* to analyse our results with higher sensitivity and temporal resolution. Similarities between these two analyses can be interpreted as mutual support of the results. Any differences between these analyses might be due to the incorporation of additional variance and temporal information in the GLMM, and, when compared with nontemporal analyses, might be indicative of the strength of the effect (i.e., weak effects need higher resolution).

Outline

In this dissertation, I present my work on how word and action learning are entangled in early childhood. In Chapter 2, we investigated whether word and action learning develop in synchrony in early childhood or whether children rather learn in one of the two domains (Eiteljoerge, Adam, Elsner, & Mani, 2019b). In Chapter 3, we extended the first study to see how preferences for one domain can influence learning in this domain. This chapter is available as a preprint on OSF (Eiteljoerge, Adam, Elsner, & Mani, in prep.a). In Chapter 4, we investigated how action consistency can influence young children's word learning: Do children learn the name of an object more easily if the object always performs the same action or always different actions (Eiteljoerge, Adam, Elsner, & Mani, 2019a)? In the Discussion, I will summarise the results and discuss them in the light of the current literature on word and action learning.

Word-object and action-object association learning across early development

Abstract

Successful communication often involves comprehension of both spoken language and observed actions with and without objects. Even very young infants can learn associations between actions and objects as well as between words and objects. However, in daily life, children are usually confronted with both kinds of input simultaneously. Choosing the critical information to attend to in such situations might help children structure the input, and thereby, allow for successful learning.

In the current study, we therefore, investigated the developmental time course of children's and adults' word and action learning when given the opportunity to learn both word-object and action-object associations for the same object. All participants went through a learning phase and a test phase. In the learning phase, they were presented with two novel objects which were associated with a distinct novel name (e.g., "Look, a Tanu") and a distinct novel action (e.g., moving up and down while tilting sideways). In the test phase, participants were presented with both objects on screen in a baseline phase, then either heard one of the two labels or saw one of the two actions in a prime phase, and then saw the two objects again on screen in a recognition phase. Throughout the trial, participants' target looking was recorded to investigate whether participants looked at the target object upon hearing its label or seeing its action, and thus, would show learning of the word-object and action-object associations. Growth curve analyses revealed that 12-month-olds showed modest learning of action-object associations, 36-month-olds learned word-object associations, and adults learned word-object and action-object associations.

These results highlight how children attend to the different information types from the two modalities through which communication is addressed to them. Over

time, with increased exposure to systematic word-object mappings, children attend less to action-object mappings, with the latter potentially being mediated by word-object learning even in adulthood. Thus, choosing between different kinds of input that may be more relevant in their rich environment encompassing different modalities might help learning at different points in development.

Introduction

The fetus can hear the rhythm and intonation of voices in the immediate surrounding (May, Byers-Heinlein, Gervain, & Werker, 2011), might feel the mother move, and moves by itself (Verbruggen et al., 2018), thereby experiencing language and movement during the latter period of gestation. As soon as the infant enters this world, she is confronted with a complex, multimodal environment encompassing both of these information types (Gogate et al., 2000, Gogate & Hollich, 2010). This rich environment can help familiarize the child with the conceptual and social aspects of their world where successful communication encompasses both the comprehension of spoken language and observed actions. While infants seem to show similar patterns of learning for word-object and word-action mappings (Gogate & Maganti, 2017), word-object and action-object mappings have only been examined in older age groups (e.g., 30-month-olds; Childers & Tomasello, 2003). In the current study, we examine children's learning of word-object and action-object associations throughout infancy and early childhood. Our aim is to compare learning from these different domains, since information from these domains are present in the input that is provided to the child from early on and at least early learning of such associations can be captured by comparably simple associative learning.

Much research has shown how early and quickly children acquire language, understanding a few words long before their first birthday around six months of age (e.g., Benedict, 1979, Bergelson & Swingley, 2012, Tincoff & Jusczyk, 2011). At a similarly young age, infants seem to perceive actions and can remember them after

seven weeks (Bahrick et al., 2002), although there is some debate as to the extent to which young infants are able to discriminate between matching and mismatching action-object associations (e.g., a woman using a hairbrush to brush her hair vs. her teeth at 5.5 months Bahrick et al., 2002 and at 6 months, Hunnius & Bekkering, 2010).

Auditory and visual information often co-occur in the child's learning environment: For example, Gogate et al. (2000) showed that 76 % of play interactions contained synchronous, temporally aligned bimodal object namings during the prelexical phase (between 5 and 9 months) with this number dropping to 59 % in the early (9 - 17 months) and 36 % in the advanced lexical (21 - 30 months) stage, with potential similarities and differences in multimodal motherese across different cultures (Gogate, Maganti, & Bahrick, 2015). This mirrors the close relationship of words and actions with objects, and shows how omnipresent their co-occurrence is, especially during the first year of life. Against this background, we examine whether the acquisition of word-object and action-object associations proceeds at a similar time scale in early childhood?

In the following, we will review the literature to-date on the learning of word-object and action-object associations. Thus far, there has been little emphasis on examining the similarity of the trajectory of learning of these two kinds of associations in early development. The finding of similar trajectories in different domains would speak to the debate about potentially shared underlying resources driving learning in both domains. Note that, given the focus of the current study, we focus here exclusively on the learning of word-object and action-object associations, and do not include the extensive literature on the learning of word-action associations, i.e., verb learning, although we note it is possible that both word-object and action-object associations are likely to be impacted by verb learning. Note also that we predominantly refer to the literature of action-object learning and not

function-object learning given that the actions in this study were not affordances or functions of the objects themselves.

The general consensus in the literature to-date appears to be that, at later ages, from around 3 to 4 years, there are few differences in children's ability to fast-map novel word-object and action-object associations (Childers & Tomasello, 2003, Dysart et al., 2016, Riggs et al., 2015). Riggs et al. (2015) showed that a) 3- to 4-year-old children choose a novel object relative to a familiar object when presented with a novel action or a novel word, b) retain the action-object association and extend it to other members of the object category as they do with words, and c) distinguish between a previously familiarized novel-action-object mapping and an unfamiliar novel-action-object mapping, showing that they had formed a detailed representation of the novel action presented. In a more recent study, Dysart et al. (2016) familiarized 3- and 4-year-olds with two novel objects. In the following test phase, the two - now familiar - objects were presented alongside another novel object ("super-novel"), accompanied by a novel label or action. In both conditions, children chose the super-novel object, suggesting that novelty drives referent selection in both word-object and action-object mapping.

Directly contrasting learning of word-object and action-object mappings, Childers and Tomasello (2003) taught 30-month-old children novel word-object and action-object mappings and found reliable learning of action-object mappings from early on, as well as extension of learning action mappings to other members of the object category. Taken together, these studies suggest that word-object and action-object association learning overlap considerably and might, therefore, develop in synchrony.

However, other studies have shown that action-object representations may indeed be prioritized relative to word-object mappings. One reason for this, suggested by Yoon, Heinke, and Humphreys (2002), is that action representations can be accessed merely via a visual route in contrast to labels that require additional linguistic information (i.e., semantics) for successfully mapping the representation with the

object. Work with adults finds that adults are faster to reach action decisions on objects (what kind of action can be performed on an object) relative to words, which is interpreted as suggesting that access to action knowledge about objects is facilitated compared to access to semantic knowledge (Chainay & Humphreys, 2002). Hahn and Gershkoff-Stowe (2010), testing 2- and 3-year-old children and adults, found that adults and young children retrieve newly learned action-object representations easier than word-object representations. Further, Dysart (2018) found that the presence of known competitors (e.g., hairbrush) supported the retention of action-object mappings after 7 days but not the retention of word-object mappings at 50 months of age.

However, in contrast to some of the action literature reported above, Puccini and Liszkowski (2012) report that, at 15 months of age, children only learned the mapping between a word and object, and not an action and an object, and only when the word was presented without an accompanying referential action. This might suggest that action-object mappings may not be as robust as word-object mappings due to potentially greater variability in the former relative to the latter (i.e., a ball can roll, fly, and bounce, but is usually only called a "ball"). This finding, along with numerous studies of word-object mapping in early childhood (see Waxman & Gelman, 2009, for a review) may be taken to suggest that early in life, children may tend to prioritize the mapping of words to objects and that auditory input dominates the scene in multimodal naming (Gogate, Prince, & Matatyaho, 2009, Puccini & Liszkowski, 2012, Robinson & Sloutsky, 2004). However, given the contrasting results from the action learning literature, the picture of learning words and actions for objects still appears to be diffuse and requires systematic investigation.

Taken together, while infants are able to recognize previously experienced associations between actions and objects and words and objects, in a multimodal environment, encompassing both of these information types, the complexity of the

input might lead to children prioritizing one information type over the other in mapping information onto objects. Accordingly, children might learn either word-object- or action-object mappings at different stages in early development, and only later be able to integrate both mappings into their object knowledge. Alternatively, if children learn about both information types already from early on, this would support the idea of a similar developmental trajectory of word-object and action-object association learning in early childhood.

Current study. Against this background, the current study examined the extent to which participants learn to associate words with objects relative to actions with objects, when given the opportunity to learn both. Thus, we presented participants with distinct action-object (Object A-Action A/Object B-Action B) and word-object mappings (Object A-Word A/Object B-Action B) for novel objects and examined differences in their learning of these at the different ages tested. In a first learning phase, participants saw two novel objects and heard distinct labels for these objects and saw distinct actions being performed on each of the objects in a temporally asynchronous manner. In a following test phase, they first saw static pictures of both objects presented simultaneously on-screen. The objects then disappeared from the screen at which point participants either heard the label in a carrier noun phrase or saw the action associated with one of the objects. Subsequently, participants saw both static pictures of the objects on-screen again and we examined their fixations to the target object as an index of their learning of the word-object and action-object association. We tested children at 12 months, 24 months, 36 months of age, as well as adults, to capture the early developmental trajectory of word-object and action-object learning.

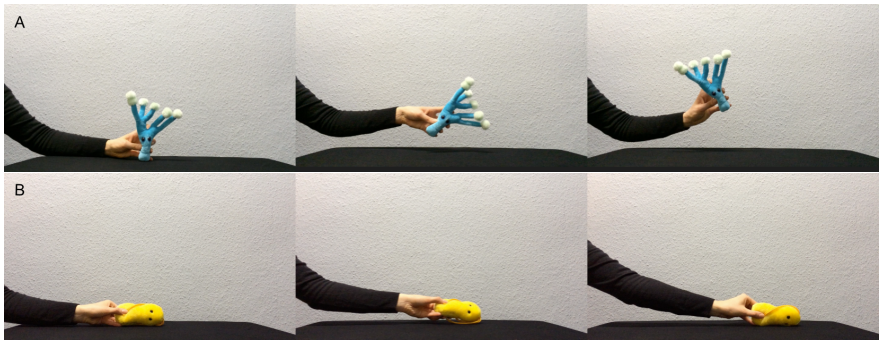
Based on the literature on the salience of word-object associations in early childhood (e.g., Fulkerson & Waxman, 2007), we predicted that 12-month-olds may show improved learning of the word-object associations relative to the action-object associations. Furthermore, we predicted that this word benefit ought to increase

with age, especially around the 24-month mark, with children on the other side of the vocabulary spurt, who may show increased sensitivity to the associations between words and objects. By 36 months of age, owing to increased cognitive resources, we anticipated that children (and indeed adults) learn both word-object and action-object associations with similar ease. We made the decision to include the 36-month-olds based on the results of the 12- and 24-month-olds.

Methods

Participants. Thirty-four German monolingual 12-month-olds (range = 11;44 - 12;59 months; mean = 12;02, female = 14), thirty-six 24-month-olds (range = 22;39 - 25;78 months; mean = 24;19, female = 17), thirty-nine 36-month-olds (range = 34;23 - 37;84 months; mean = 35;84, female = 17) and 27 adults (range = 19;06 - 28;71 years; mean = 22;44, female = 19) participated in the experiment. Fourty-eight additional participants were tested but excluded from the analysis because of piloting (5), unwillingness to participate (2), technical failure (5), own cat was called "Loki" which resembles one of the novel words (1), bilingualism (3), preterm birth (1), or insufficient data (31, see Preprocessing). Thus, exclusion was overall at 26.1 %, with 16.9 % for data reasons and 11.5 % for any other reasons. Children were recruited from the babylab database and participation was rewarded with a book. Caregivers of 12-month-olds completed the ELFRA-1 (Elternfragebögen für die Früherkennung von Risikokindern; Grimm & Doil, 2006), while caregivers of the 24-month-olds completed the German adaption of the FRAKIS (Fragebogen zur frühkindlichen Sprachentwicklung; Szagun, Stumper, & Schramm, 2009) in order to control for children's language abilities. Additionally, cognitive and fine motor tests using the Bayley Scales of Infant and Toddler Development were administered for both age groups in the lab (Bayley, 2006). Parents signed a written informed consent form for their child. Adults were mostly students of the University, also signed the written consent form, and were rewarded with either 0.5 course credit

Figure 2. Blue and yellow germ toys were used as novel objects. As novel actions, an upward movement with tilting to the sides (Panel A), and a sideways movement with tilting backwards and forwards (Panel B) were used.



points or 4 €. Ethics approval was granted by the ethics committee of the Georg-Elias-Müller-Institute for Psychology, University of Goettingen (Project 123).

Stimuli. We selected two novel words in keeping with the phonotactic rules of German (Tanu and Löki), two arbitrary actions (described in further detail below), and two novel objects (a yellow and a blue soft toy germ from `https://www.giantmicrobes.com/us/`, see Fig 2). Audio stimuli were recorded by a female German native speaker in infant-directed speech. The labels were embedded in separate carrier phrases for the training (e.g., "Schau mal, ein Tanu!" meaning "Look, a Tanu!") and test phase (e.g., "Hey, guck mal, wo ist denn das Tanu?", meaning "Hey, look, where is the Tanu?").

Training phase stimuli. Training phase stimuli presented children with both word-object and action-object associations during the course of each trial. The visual stimuli presented a hand (with the arm of the agent being visible) moving one of the objects according to two selected actions. When moving up and down (Fig 2, panel A), the hand began on the lower middle of the screen and slowly moved upwards and then down again while slightly leaning to the right and to the left in an alternating manner. When moving sideways (Fig 2, panel B), the action began in the middle of the screen and the object would slightly turn backwards and forwards while moving first to one side of the screen, then backwards to the other, and ending up in the middle of the screen again, and both actions were roughly

7 s long. We chose these actions in preference to some previously selected actions with younger infants (e.g., shaking and looming) since these previously selected actions might be used to draw attention to the object and need not necessarily be considered object-related actions.

Auditory stimuli were presented before the onset of the action in half of the trials, and after the action had ended in the other half of the trials to control for influences of order of presentation of word-object and action-object associations. Each video in the training phase was ten seconds long, 720×420 pixels in size, and presented in the middle of the screen.

Test phase stimuli. The stimuli for the test phase included baseline stimuli, prime stimuli and recognition stimuli. Baseline and recognition stimuli were identical and involved children being presented with stationery images of the two objects in silence. Images (baseline and recognition) were 640×480 pixels and appeared next to each other in the center of the left and right half of the screen. Prime stimuli involved children hearing either the label for one of the objects or videos of the action movements described above, although the actions were now performed in the absence of the object. This was to ensure that the action-object association was not prioritized in any way and kept as similar to the word-object prime phase as possible (including the absence of a referent). Seeing the action being performed on the object would cue children to the identity of the correct object mapping. We also wanted to avoid using a novel control object because it could be considered as a new reference object. The videos of the action movements lasted seven seconds and were presented in the center of the screen. The label for the target object was embedded in a carrier phrase that similarly lasted seven seconds, with the label occurring at the end of the phrase.

Procedure. Participants sat in a dimly lit and quiet experimental room at a distance of 65 cm from a TV screen (92×50 cm), where the visual stimuli were presented. A remote eye tracker (Tobii X 120), set on a platform underneath the

TV screen, was used to record gaze data at 60 Hz. Stimuli were presented using E-Prime software. Auditory stimuli were presented through two loudspeakers situated above the television screen. Further, two video cameras centered above the screen served to monitor the participant during the experiment. Prior to testing, gaze was calibrated using a 5-point calibration procedure, in which a red point appeared in every corner and the center of the screen. The experiment only started following successful calibration. Each trial began with a Teletubby serving as a fixation getter in the middle of the screen against a black background, followed by the stimulus presentation.

After the eye-tracking part, the cognitive and finemotor scales of the Bayley scale were administered with the 12- and 24-month-olds (Bayley, 2006). This was part of a larger cross-laboratory project to examine the correlation between performance in these scales across development. These data and the vocabulary data will not be reported here.

Experimental design. Each child was presented with a training phase and a test phase (see Tab 2). We ensured that word-object and action-object associations were counterbalanced across children resulting in four different training lists with each of the objects being associated with each of the labels and each of the two actions, and according test phases. Presentation order of all trials within the individual training and test phases was fully randomized.

Training phase. This phase consisted of eight trials where children saw individual objects onscreen with four trials per object. In each trial, children were presented with one object and heard a distinct label for this object and saw a distinct action being performed on the object. Each trial presented both the word-object and the action-object association, with half of the trials presenting the word-object association first and the other half of the trials presenting the action-object association first. This is in contrast to studies that ensured temporal synchronicity between word and action presentations (e.g., Gogate & Bahrick, 1998, Matatyaho-Bullaro

et al., 2014). In our study, we deliberately avoided temporal synchronicity for two reasons. First, this would help to ensure that children did not associate the word with the action (similar to the literature on word-action learning, i.e., verb learning). Second, we wanted to give children time to process each association separately, which has been shown to support learning (Sloutsky & Napolitano, 2003).

Test phase. Each test trial was divided into a baseline phase, a prime phase and a recognition phase. During the baseline phase, children saw both static objects side-by-side on screen in silence for 2.5 s. The objects disappeared after 2.5 s. Next, in the prime phase of 7 seconds (similar to Von Holzen & Mani, 2012), children either heard the label for one of the objects (word-object trial) or saw a hand performing one of the actions (action-object trial), in the absence of the object. The recognition phase began 300 ms after the offset of the label or end of the action, at which point the objects reappeared on-screen as static pictures once again for 2.5 s in the same positions as in the baseline phase. Across eight test trials, children were tested equally often on their knowledge of the word-object and action-object associations (four trials each) for both objects.

Following these eight critical trials, we examined whether participants had associated the words with the actions presented (word-action trial) across four trials. We included these control trials to examine the possibility that participants had formed a word-action association that may influence responding in our critical test of word-object and action-object association learning. Here, the actions were presented side-by-side in silence (baseline phase) for 6500 ms. Then, in a prime phase, participants were presented with a black screen and heard one of the labels. 300 ms after the offset of the label. During the recognition phase, the actions reappeared on-screen in the same positions for further 6500 ms.

Note that both word-object and action-object association types were tested as a within-subject factor, and we refer to these as “conditions” in the rest of the paper.

Table 2
Overview of Learning and Test phase. The order of trials within one phase was fully randomized.

Phase	N trials	Description
Learning phase	8	2 novel objects (Ãă 4 trials) with distinct novel label and novel action
Test phase (critical)	8	4 word-object test trials and 4 action-object test trials
Test phase (control)	4	word-action test trials

Pre-processing. The eye-tracker provided an estimate of where participants were fixating in each time stamp during the trial, with one data point approximately every 16 ms. All data (gaze data and trial information) were exported from E-Prime and then further processed in R (R version 3.2.4 (2016-03-10), R Core Team, 2016). For each time stamp, data were only included when one or both eyes of the participant were tracked reliably (validity less than 2 on E-Prime scale). When both eyes were tracked, the mean gaze point for both eyes was computed for further analysis. Gaze data were then aggregated into 40 ms bins. Two areas of interest covering the location and size of the two images presented on-screen respectively were defined based on the location and size of the two images on-screen in the test phase.

For the baseline and recognition phase, we coded whether the participant looked at the correct object (i.e., the target), the distractor or at neither of these onscreen for each timepoint. Typically, fixation data is analyzed from 240 ms to 2000 ms from the onset of a label (Swingley, Pinto, & Fernald, 1999), since only these fixations can be reliably construed as a response to an auditory stimulus. Following this convention, we analyzed only data from 240 ms from the onset of the baseline phase and the recognition phase. This ensured that we avoided early fluctuations in fixations as a result of the images appearing on-screen. Note that this is in addition to the 300 ms between the offset of the label and the action at the end of the prime phase, thereby allowing participants adequate time to process the information provided.

Across word-object, action-object and word-action trials, we corrected the recognition phase for any fixations to either object in the baseline phase on an individual trial level (i.e., each time point in the recognition phase was corrected for the overall looking score in the baseline phase on that particular trial). This baseline-corrected measure was our dependent variable in all analyses.

In order to ensure that we analyzed similar durations of the word-action and word-object and action-object trials, we only analyzed 2000 ms of the recognition phase of the word-action trials. Since we wanted to use target looking in the word-action condition as a control variable in analyzing participants' learning of word-object and action-object associations, we calculated the difference in preferential target looking from the baseline to the recognition phase in word-action trials and included this value as an index of the formation of word-action associations as a control variable.

Single test trials were excluded if a participant was looking at the stimuli less than M − 3SD of the time. Thus, a trial was excluded if a 12-month-old child looked less than 8.4 % of the time during the trial to one of the two pictures. This led to an exclusion of 27 trials (5.24 %). For 24-month-olds, this criterion was at $\leq$ 30.3 %, 26 trials (5.67 %); for 36-month-olds it was at $\leq$ 34 %, 34 trials (6.35 %); and for adults at $\leq$ 50 %, 11 trials (3.13 %). This led to one further exclusion as one participant did not look enough during any of their trials (24: N = 1). Furthermore, participants were excluded from the analyses if the participant contributed only one trial per condition in the test phase (12: N = 8; 24: N = 5; 36: N = 5; Adults: N = 3). Further, nine participants were missing data in either the baseline or the recognition phase in one of the conditions (12: N = 5, 36: N = 3, Adults: N = 1). Together, this led to 13 12-month-olds, six 24-month-olds, eight 36-month-olds, and four adults, and left us with 136 included participants in the final sample.

Data analysis

Within each age group, we report two statistical approaches here. First, we report traditional t-tests and correlations. In particular, we ran a paired samples t-test comparing the proportion of target looking in the word-object and the action-object trials. Then we ran separate one sample t-tests comparing the baseline-corrected proportion of fixations to the target during the recognition phase against chance (0 = no difference between baseline and recognition phase). Next, we report generalized linear mixed models. These models have two advantages over traditional analyses. First, they allow us to control for variability along additional dimensions, such as object, label, action, and performance in word-action trials. Second, they also allow us to include the factor Time during the trial, thereby examining looking patterns over the whole time course of the trial (for a more detailed description and instruction see Mirman, 2014, Mirman, Dixon, & Magnuson, 2008).

We included Time and its second, third, and fourth polynomial in the model. This allows us to model our data as a linear, quadratic, cubic, and quartic function of time. In Fig 1, we plotted theoretically expected and typically observed curvatures of target recognition. If participants did not learn the tested associations, and therefore, do not show target recognition at test, they typically look back and forth between the two objects, meaning that their target looking is around chance level (pictured in light blue) throughout the trial. If participants have learned the association, and therefore show target recognition at test, their target looking would be first at chance level, would later increase when they look to the correct object, and goes back to chance level (pictured in dark blue). This would be reflected in a quadratic curve, or a quartic curve when tails are included at the beginning and end. In practice, adults usually look at the correct object until the end of the trial when they have learned the correct association (pictured in orange), thereby showing their recognition of the target (Mirman et al., 2008). This curve is often better fitted with a linear (steep rise) or cubic function (steep rise and high plateau).

Children's responses are often weaker and increases in target looking might start slightly later (pictured in yellow). However, their responses are also typically more variable requiring greater flexibility in model fitting and interpretation.

We fitted a generalized mixed model (GLMM) using lme4's lmer function in R (Bates, Mächler, Bolker, & Walker, 2015) with Gaussian error structure and identity link function. Condition and Time, and their interaction were included as fixed effects of interest. Further, Object, Name, and Action, as well as the baseline-corrected proportion of target looking in the word-action condition were included as fixed effects. We included performance in the word-action control trials in the model to ensure that performance in the word-object and action-object conditions was not driven by spurious associations being formed between words and actions. We also included Participant id and Condition as random factors to allow for random slopes across participants. The reduced model did not include Condition. A comparison between the reduced model and the full model including the factor Condition then allowed us to evaluate the influence of the factor Condition (action-object associations vs. word-object associations). We used the function drop1 to evaluate the influence of each factor in the model. This function compares the model including one factor with a model without this factor, and thereby evaluates its contribution to the model.

Visual inspection of a qq-plot and a histogram of the residuals showed a normal distribution, but, as is usually the case with looking time data, homogeneity seemed to be violated when plotting residuals against fitted values. Log-transforming the response did not contribute to an improvement of the model. Therefore, we chose the first model but results need to be reviewed with care.

For ease of interpretation and due to the complex interactions between conditions across the four age groups found in an omnibus ANOVA and GLMM, we do not report these analyses here. The results of these analyses can be found on `osf.io/`

`b7fmt`. Here, we report only analyses split by age group to better capture the learning effects within each age group.

Results

Descriptives for each condition per age group can be seen in Table 3. An overview of the results can be seen in Table 7 and Table 8, and in Fig 3 and Fig 4.

Table 3
Descriptives of baseline-corrected proportional target looking in the recognition phase of the word-object and the action-object condition per age group (12, 24, 36 months, and adults). Values higher than 0 indicate target looking, values below 0 indicate distractor looking.

Age group	Condition			
	word-object		action-object	
	Mean	SD	Mean	SD
12-month-olds	.01	.16	.02	.13
24-month-olds	.03	.17	-.004	.15
36-month-olds	.1	.15	.003	.1
Adults	.26	.29	.25	.19

12-month-olds.

T-tests. For the 12-month-olds, a paired samples t-test comparing word-object and action-object conditions found no significant difference between conditions, $p = .52$. Separate one sample t-tests comparing baseline-corrected target fixations in each of the conditions to chance (chance = 0) were not significant (all ps $>.1$).

GLMM. For the 12-month-olds, the model comparison between the full model including condition (word-object vs. action-object) and the reduced model was significant ($\chi^2 = 15.31$, df $= 5$, $p = .009$). Using drop1, the model revealed a significant interaction of Condition*poly3 ($\chi^2 = 5.58$, df $= 1$, $p = .018$). The interaction of Condition*poly2 was at $\chi^2 = 3.30$, df $= 1$, $p = .069$. Thus, the model revealed differences mainly in the cubic curvature across conditions. This captures differences in two peaks in the time course of fixations across the two conditions, hence necessitating the cubic curve. As Fig. 3 shows, fixations in the word-object

condition show a slight bump early in the trial flattening out towards the end of the trial, while fixations in the action-object condition show a later bump from around 500 ms with an increase in fixations to the target in this condition relative to the word-object condition.

For the model, examining the data of the word-object trials alone (word-object split model), none of the time terms were significant. The influence of performance in the word-action control trials on target fixations in the word-object condition was at $\chi^2 = 3.8$, df = 1, $p = .051$ (a simple correlation confirmed this effect with performance in the word-object condition positively correlating with performance in the word-action trials, rho (Spearman) = .4, p = .018).

For the model, examining the data of the action-object trials alone (action-object split model), the quadratic time term ($\chi^2 = 4.89$, df = 1, $p = .027$), and the quartic term were significant ($\chi^2 = 6.61$, df = 1, $p = .01$). As indicated in Fig 3, 12-month-olds showed an increase in target fixations around 500 ms into the trial, captured by the quadratic curve, while the quartic curve captures two changes in the curve of fixations, with an early slight dip in target looking at the beginning of the trial and a later stronger increase to the target around 500 ms. Such a pattern of looking is in keeping with our theoretical expectations of target looking behaviour, with a classic quadratic curve when only looking at the later part of the trial and a slower increase towards target looking early in the trial.

In summary, for 12-month-olds, , we find evidence for differences in looking behaviour in the test phase (from the baseline phase) in the action-object condition. Such a pattern of looking behaviour is typically been interpreted as evidence of target recognition (see our discussion of expected patterns above). These differences in target fixations were not significant across the entire time course of the trial, as indicated by the t-tests reported, highlighting potential concerns with the robustness of this effect at this age.

24-month-olds.

T-tests. For the 24-month-olds, a paired samples t-test comparing word-object and action-object conditions found no significant difference between conditions, p = .37. Separate one sample t-tests comparing baseline-corrected target fixations in each of the conditions to chance (chance = 0) were not significant (all ps >.1).

GLMM. For the 24-month-olds, the model comparison between the full model including condition (word-object vs. action-object) and the reduced model was not significant (χ^2 = 2.83, df = 5, p = .73). Using drop1, the model revealed no interaction between Condition and any of the time terms.

To summarize, at 24-months, we did not find robust evidence for action-object learning nor for word-object learning. However, and we did not find a difference in target recognition across conditions (in both analyses, t-tests and GLMM)[1].

36-month-olds.

T-tests. For the 36-month-olds, a paired samples t-test revealed significant differences between word-object and action-object conditions (t(38) = -3.39, p = .002, d = -0.77). Separate one sample t-tests comparing baseline-corrected target fixations in each of the conditions to chance (chance = 0) found a significant difference from chance in the word-object (t(38) = 4.27, p <.001, d = 0.97) but not the action-object condition (t(38) = 0.12, p = .9, d = 0.03). These results suggest that 36-month-olds learned word associations but not action associations for objects.

GLMM. For the 36-month-olds, the model comparison between the full model including Condition (word-object vs. action-object) and the reduced model was

[1]For consistency with the other age groups, we split the data from the word-object and action-object trials. For the word-object split model, the cubic time term was significant (χ^2 = 4.04, df = 1, p = .044), suggesting that 24-month-olds showed increased looks to the target object (followed by a decline) relative to baseline. For the action-object split model, the linear time term was significant (χ^2 = 4.11, df = 1, p = .043) which reflected a slow decrease in target looking over time (reduced relative to baseline) and does not match theoretical expectations with regard to the curvature associated with target recognition. Given issues with the interpretation of this result against the background of non-significant model comparison, we report these results here for thoroughness but will not interpret this as evidence of learning until further data corroborates this finding.

significant (χ^2 = 36.34, df = 5, p <.001). Using drop1, the model revealed a significant interaction of Condition and the cubic time term (χ^2 = 20.59, df = 1, p <.001), suggesting that there were differences across conditions.

For the word-object split model, the cubic time term was significant (χ^2 = 6.68, df = 1, p = .01). Also for the action-object split model, the cubic time term was significant (χ^2 = 15.20, df = 1, p <.001). As is evident in Fig 3, these results suggest that 36 month-olds fixate the target consistently across the recognition phase of the trial (more than at baseline) in word-object trials. In contrast, in action-object trials, they display a brief period of initial target recognition fixating the target above chance (increase from baseline) in the recognition phase, followed by fixations towards the distractor and again towards the target.

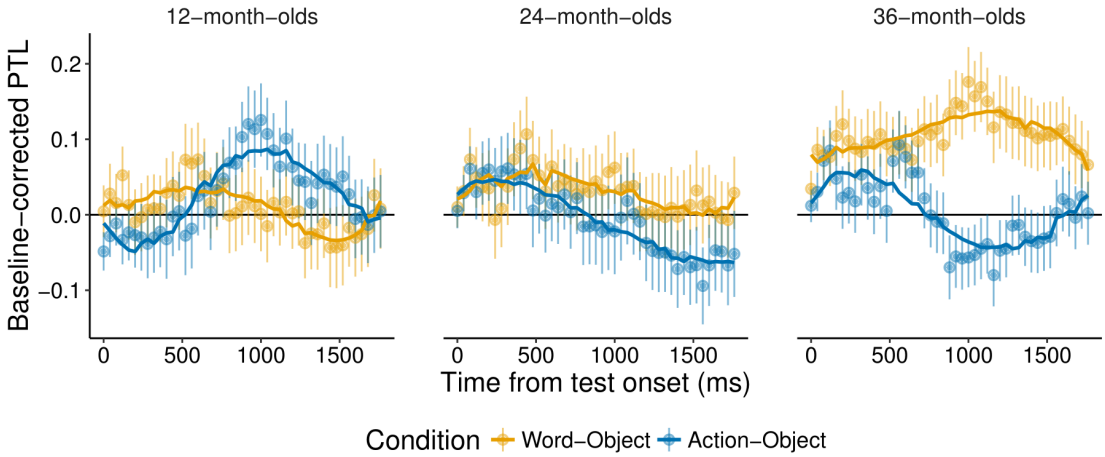
To summarize, 36-month-olds showed systematic looking patterns in both the word-object and the action-object condition, although the effects of target recognition were more in keeping with theoretical expectations for the word-object condition compared to the action-object condition. Taken together with the results of the t-tests, we interpret these results as being indicative of word-object association learning but not action-object association learning at 36 months.

Adults.

T-tests. For the adults, a paired samples t-test comparing word-object and action-object conditions found no significant difference between conditions, p = .83. Separate one sample t-tests comparing baseline-corrected target fixations in each of the conditions to chance (chance = 0) revealed significant differences from chance in both conditions (word-object: t(26) = 4.85, p <.001, d = 1.32; action-object: t(26) = 6.91, p <.001, d = 1.88), suggesting that they formed both word-object and action-object associations.

GLMM. For the adults, the model comparison between the full model including condition (word-object vs. action-object) and the reduced model was not significant (χ^2 = 4.72, df = 5, p = .45). Using drop1, the model revealed no sig-

Figure 3. Time course graphs for each of the children's age groups of baseline-corrected preferential target looking (PTL) during the recognition phase. The word-object condition is represented in yellow, and the action-object condition in blue. The line at 0 represents chance level (increase from baseline), fixations above the line indicate target looking whereas fixations below the line indicate distractor looking. The first 240 ms are cut to allow for fixation time, and the time within the trial has been corrected, so that the x-axis starts at 0.

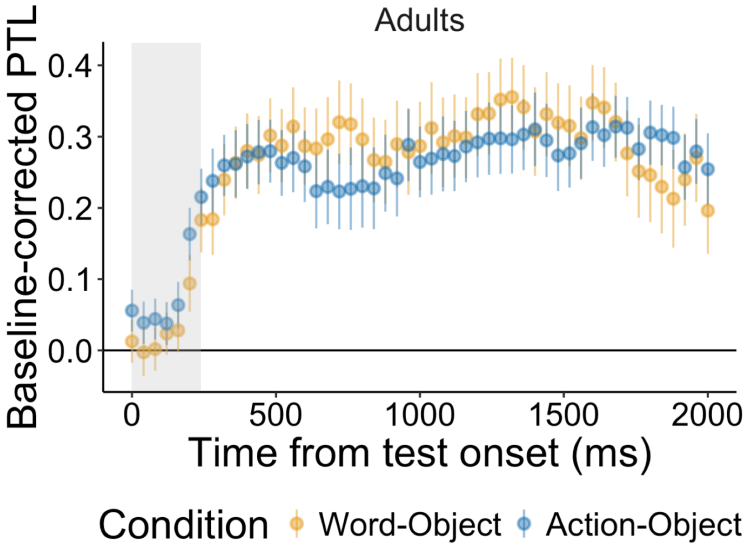


nificant interaction of Condition with any of the time terms. Thus, there were no significant differences between conditions.

For the word-object split model, the quadratic term (χ^2 = 5.54, df = 1, p = .019) and the quartic time term were significant (χ^2 = 3.86, df = 1, p = .049). Furthermore, there was an influence of performance in the word-action condition on looking times in the word-object condition (χ^2 = 7.07, df = 1, p = .008). For the action-object split model, the cubic time term was significant (χ^2 = 4.55, df = 1, p = .033). Furthermore, there was a significant influence of performance in the word-action condition on looking times in the action-object condition (χ^2 = 9.29, df = 1, p = .002), confirmed in a correlation analysis, (rho (Spearman) = .58, p = .002). Importantly, both analyses controlled for performance in the word-action control trials when examining the effect of learning of word-object and action-object associations.

To summarize, adults learned the associations in both word-object and action-object conditions.

Figure 4. Time course graphs for the adults' baseline-corrected preferential target looking (PTL) during the recognition phase. The word-object condition is represented in yellow, and the action-object condition in blue. The line at 0 represents chance level (increase from baseline), fixations above the line indicate target looking whereas fixations below the line indicate distractor looking. Note that adults' target looking starts already at the beginning of the trial. Nonetheless, we discarded the first 240 ms of the trial to match the children's analyses.



Discussion

In the current study, we explored the developmental trajectory of children learning to associate words and actions with objects. In particular, we asked, whether there are differences across development and across information types in the learning of such associations when children are given the opportunity to learn both a word and an action association for an object. In a first training phase, participants were familiarized with novel objects that were associated with both a novel label and a novel action. In the following test phase, participants' learning of word-object and action-object associations was tested by measuring their looks to the target object following presentation of either the associated action or the associated word. 12-month-olds showed some evidence of learning of action-object associations but not word-object associations when considering time within the trial, 24-month-olds learned neither word-object nor action-object associations (and there was no evidence for a difference in performance across conditions), 36-month-olds also showed

learning of word-object associations but no systematic learning of action-object associations, and adults learned both word-object and action-object associations. These results plot the developmental trajectory in children and adults' learning of word-object and action-object associations.

Our results are in keeping with the developmental pattern that has been observed in the related field of iconic gestural learning in typically developing children. Namy, Campbell, and Tomasello compared children's learning of iconic (a hopping rabbit) and arbitrary (side-to-side motion) gestures across early childhood and found that 18-month-olds and 4-year-olds successfully mapped both types of gestures with the associated objects, while 26-month-olds only learned to map iconic but not arbitrary gestures to the associated objects (Namy, Campbell, & Tomasello, 2004, see also Namy & Waxman, 1998). The authors argue that this U-shaped pattern emerges due to a reorganization of the gestural system at around two years of age, potentially due to an increased focus on linguistic communication from this age.

Our study extends these findings from iconic gestural learning to compare children's learning of word-object and action-object associations across this malleable period of development. Our findings in action-object association learning are consistent with those reported by Namy and colleagues. 12-month-olds, like the 18-month-olds in Namy et al. (2004), learnt only action-object associations, while 24-month-olds and 36-month-olds did not. Adults, like the 4-year-olds in Namy et al. (2004), learned both word and action associations for objects, with no difference between the two conditions, suggesting a period of further flexibility in learning associations for objects past the age tested in Namy et al. (2004). Thus, as far as action-object associations are concerned, our results, together with Namy et al. (2004), highlight a developmental pattern in early gesture-object and action-object association learning. Before exploring potential explanations for this developmental curve, we next discuss the results for the word-object condition in detail.

The results from the word-object condition differed considerably from the action-object condition, with younger children showing no evidence for word-object association learning. However, the literature on word learning suggests that infants understand words for objects (such as bodyparts) already at 6 months (Tincoff & Jusczyk, 2011, Tincoff, Seidl, Buckley, Wojcik, & Cristia, 2018). Further, when presented with a novel word for a novel object, children as young as 6 to 12 months of age learn to associate words with objects (Fulkerson & Waxman, 2007, but see Taxitari, Twomey, Westermann, & Mani, under review, that such learning is contingent on a number of factors, e.g., similarity of the two objects presented). Those studies, however, only presented infants with word-object associations and not both word- and action-object associations.

Studies presenting children with asynchronous input including multiple information types (as in the current study) indeed show different results. With younger children, studies presenting temporally synchronous multimodal input suggest that the temporal alignment of word and action may bolster word-object association learning at very early ages (6 - 8 months, Matatyaho & Gogate, 2008). Relatedly, studies presenting children with word associations for objects in motion, suggest that actions that are simultaneously presented, albeit not temporally aligned, with label information might help to highlight the connection between the word and the object, thereby boosting learning of word-object associations (Werker et al., 1998). Nevertheless, Puccini and Liszkowski (2012) find that young infants (at 15 months of age) learned only word associations for objects (and not gesture-object associations), and only when word-object associations are presented without gestural information, potentially due to such gestures drawing attentional resources away from learning of word-object associations. Thus, even 15-month-olds' word learning is impacted by the presentation of different information types, with them learning word-object associations only when the stimuli are presented in the absence

of such gestural information, at least when information from the gestural modality is asynchronous.

With regards to the results of the 12-month-olds, we found action-object association learning when we considered time within the trial, but not when we collapsed the data across time. This suggests, that learning in this condition could only be observed with high resolution, and therefore, might still be weak at this age. Nonetheless, the significant model comparison suggests that action-object association learning differed from word-object association learning, which we did not find in any of the analyses at this age. Our results also highlight additional factors that may potentially influence word-object learning in 12-month-olds. We found a positive correlation between 12-month-olds' performance in the word-object condition and the word-action condition. In other words, those 12-month-olds who mapped the words onto the actions performed on the objects that these words referred to (the word-action condition), also learned the word-object associations better. This finding raises the possibility that word-object association may be mediated by spurious associations between words and actions at this age, further highlighting the salience of actions at this age. Furthermore, it could be that 12-month-olds do not to choose between word-action and word-object associations for the same words, even though the words and the actions were not temporally aligned. Accordingly, it could be that some 12-month-olds learned the word-object association, while at the same time maintaining action associations for this word (c.f., Roembke & McMurray, 2016), at the very least when the actions and words are not temporally synchronized. Such multiple associations might be restricted later in childhood, at two years of age, where we found no similar correlations between performance in the word-action and word-object condition, potentially due to children using morphosyntactic cues in the linguistic stimuli provided to map the words onto objects alone (and not actions; Waxman et al., 2013).

Next, we discuss the findings from the 24-month-olds in further detail. Previous studies suggest that 24-month-olds learn words after minimal exposure (see Westermann & Mani, 2017, for an extensive review of early word learning). However, Hahn and Gershkoff-Stowe found that action-object associations were learned better than word-object associations at both 2- and 3-years of age (Hahn & Gershkoff-Stowe, 2010, see also Childers & Tomasello, 2003, but see the results by Namy et al., 2004, Namy & Waxman, 1998).

Our findings reveal no evidence for either action-object association learning nor word-object association learning at 24 months (given the absence of an influence of condition on target looking in the model comparison). We note that when we look at the time course of the 24-month-olds in both conditions, we see target looking at the beginning of the trial, but this target looking decreases over time, and therefore, does not mimic theoretically expected patterns. Together with the contrasting results in the literature to-date, it appears that both the gestural and the linguistic communication system at 24 months may be more susceptible to interference from other information types, especially when this information is presented simultaneously, albeit asynchronously. Indeed, this is in keeping with recent literature highlighting the influence of additional factors, e.g., language background, on word-object association learning at the same age (Gogate & Hollich, 2016).

By 3 years of age, we find systematic target looking in both the word-object and action-object condition, but found no modulation of the effect by performance in the word-action condition, similar to the 2-year-olds. However, we note that target looking in the action-object condition of the 36-month-olds followed a cubic curve with an initial increase towards the target, followed by distractor looking. In contrast, target looking in the word-object condition showed an early increase towards the target, which was sustained throughout the course of the trial. The difference in the time courses of these effects, and also when compared to theoretical assumptions (see Fig. 1), might be taken to suggest that children did not learn action-object

associations at this age, relative to word-object learning and action-object learning at older ages, e.g., at 4 years (Dysart et al., 2016). This is further evident in the non-significant result when comparing overall target looking against chance. Thus, for slightly younger children such as our 36-month-olds, verbal information might continue to receive more attention overall relative to action information.

Adults learned both word-object and action-object associations successfully, with no difference between conditions. Further, we found that performance in the action-object condition was positively correlated with performance in the word-action condition, raising the possibility that action-object associations in adulthood may be mediated by language, i.e., via the word-action and word-object mappings. However, we note that the GLMM found an influence of performance in the word-action condition on target fixations overall (across both conditions), potentially reflecting overall association strength for all three associations presented rather than specific modulation of performance in the action-object condition by performance in the word-action condition. Importantly, the inclusion of performance in the word-action control trials in the model revealed effects of word-object and action-object association recognition over and above potential effects of spurious associations between the words and actions presented.

Overall, our results suggest some evidence of early learning of action-object associations but not word-object associations, with increased learning of word-world mappings only later in childhood, given the physical and temporal attributes of the stimuli presented in the current study. Potentially, this shift to word-world mappings may occur due to children's greater exposure to systematic one-to-one mappings between words and objects across the second and third year of life. Nevertheless, with increasing lexical exposure, children develop and might learn both word-object and action-object associations.

The early action learning effects call into question theories of language acquisition that highlight a special place for words in input-world mappings (e.g., Waxman

& Gelman, 2009), and suggest rather that early mappings may be driven more by the salience of input available, with the potentially more salient action-object mappings taking precedence early in childhood. With time and increased exposure to word-world mappings, children may begin to attend more to the mapping between words and the world around them. This is further amplified with greater linguistic and cognitive competence, such that older children and adults learn to track word and action associations for the same objects.

Finally, we include a note on our statistical analysis. Although the results from the different statistical analyses diverged in some cases, we believe that the combination of t-tests and GLMMs allowed us to quantify our effects in a differentiated way. For example, the GLMMs suggested a quadratic curvature for 12-month-olds' action learning. This pattern was not observable when target looking was averaged across time and t-tests were employed. These differences might suggest that target looking was rather brief and only observable when time within the trial was considered. Similarly, 36-month-olds' target looking in the action-object condition manifested itself in a cubic curve with its form not being predicted by the theoretical assumptions. Here as well, the combination of t-tests and GLMMs proved valuable, highlighting the short-lived nature of 36-month-olds' action-object association. Together, the combination of the different tests allowed us to better interpret our data and present a more differentiated picture of the development of early word and action learning.

Taken together with the previous literature on the topic, our findings suggest a distinct developmental pattern in children's mapping of words to objects and actions to objects. We observed that children showed a preference for action-object associations early in life (at 12 months) and only later showed learning of word-object associations (at 36 months). It was much later in development, that we found some evidence for the maintenance of both action-object and word-object associations, leading to adults learning both word-object and action-object associations. These

results advance our understanding of children's learning from action and word input in their environment: when presented with rich learning environments, selecting interesting or relevant aspects helps humans of all ages to structure the rich and multimodal input they perceive in the world.

Do preferences for words and actions influence word-object and action-object learning in early childhood?

Abstract

Children live in a multimodal world: For example, communication with young children not only includes information from the auditory linguistic modality in the form of speech but also from the visual modality in the form of actions that caregivers use in the interaction with children. Dynamic systems approaches suggest that multimodal input can help children to learn from the environment while also allowing the child to shape their own learning experience through selective attention. This selective attention might be influenced by the child's preferences, which, in turn, might shape the child's learning behaviour. In this study, we investigated how children's selective attention to information from the linguistic or the action modality influence learning in both domains.

Two- to 3-year-old children and adults participated in a novel gaze-contingent paradigm that allowed them to choose between being provided with the labels for or the actions that one can do with novel and familiar objects. At test, participants saw the two novel objects and either heard one of the labels or saw one of the actions that had been performed on one of the objects. Following label and action presentation, we investigated whether children fixated the target object, i.e., the object whose respective action/label had been presented, as an index of word and action learning. Children learned word but not action-object associations, and their target looking in the word-object condition was influenced by their selective attention to words in the earlier phase. Adults learned word-object associations and action-object associations, and their target looking in the action-object condition was influenced by their selective attention to actions in the earlier phase.

Gaze-contingent eye-tracking paradigms provide us a unique method to analyse children's active learning preferences, which will help us better understand children's

learning behaviour in a complex world. In particular, we show that in multimodal environments, children's preferences might help to structure the complex input into chunks that are compatible with the child's cognitive capacities in the moment.

Introduction

From early in life, children are exposed to a complex multimodal world. In such a rich environment, it may be helpful to filter the complex input available by selectively attending to relevant or pertinent information. However, what is perceived as relevant information may be influenced not only by salience, frequency, or other people's focus, but also by our own log- or short-term individual preferences (i.e., subjective relevance). For example, if a child is presented with a frog that jumps around while the caregiver labels the frog, the child might attend either to the action or the word in the input and associate this with the object, the frog. Only when the perceptual abilities of the child meet the requirements to attend to and learn from both modalities in this scenario, might the child be able to learn both aspects. Until then, what the child learns might be a complex interaction of the child's abilities and interests, and the immediate environment. Indeed, only once the cognitive and perceptual abilities of the child are sufficiently advanced (Eiteljoerge et al., 2019a, 2019b), might the child be able to simultaneously attend and learn from information from multiple modalities. Until then, what the child learns might be a complex interaction of the child's abilities, interests, and the immediate environment.

The idea of an active, self-exploring child has had a long history in both theoretical and educational work in psychology (Montessori, 1914, Piaget, 1936). It has been further formulated in dynamic systems theories that assume that no specific aspect of a situation is dominant per se, but it is rather the interaction between the child and her immediate environment which shapes the learning experience (Hirsh-Pasek et al., 2000, Smith & Thelen, 1993). In such a theoretical framework,

the child's interests can direct the focus of her attention, and thereby influence the learning outcome (for a discussion on other theories and their contribution to active word learning, see Mani & Ackermann, 2018).

Indeed, the idea that children are actively engaged in learning has been supported in the recent years by many studies on children's curiosity-driven learning and exploratory play. For example, Kidd and colleagues document the so-called Goldilocks effect, showing that even 7- to 8-month-olds chose what they attend to when presented with complex stimuli, avoiding both overly simple and complex scenarios with regards to their current state of knowledge (Kidd et al., 2012, 2014). Further, using head-mounted cameras to examine attention to objects and caregivers in exploratory play, Yu and colleagues suggest that the perspective of the child's egocentric view plays a crucial role in their early word learning (Smith, Jayaraman, Clerkin, & Yu, 2018, Yu & Smith, 2013). Later still, 3- to 5-year-olds appeared to learn better in an active touch screen task compared to their yoked control counterparts (e.g., Partridge, McGovern, Yung, & Kidd, 2015) and 6- to 8-year-olds showed improved memory in an object recognition game for actively compared to passively learned object locations (Ruggeri, Markant, Gureckis, & Xu, 2016). Nonetheless, this effect seems to be dependent on age (Kirkorian, Choi, & Pempek, 2016), and on other factors such as gender, socio-economic status, and the movements required to interact with a recording device (i.e., tapping or dragging, Russo-Johnson, Troseth, Duncan, & Mesghina, 2017). Relatedly, neurocomputational models show successful learning if the input and the current state of the learner are geared to one another, reflecting the changing dynamics that might be at work in infants' learning (Twomey & Westermann, 2017).

This is not, however, to suggest that the exposure to language and communicative input that the child receives in her environment is unimportant. Language is acquired knowledge — thus what babies see and hear is critical to learning. Indeed, the amount and the quality of exposure to communicative interactions that

children receive has long been shown to play a crucial role in early language acquisition, as Smolak and Weinraub (1983) suggested. In their study, mothers of highly verbal 2-year-old children produced more utterances as well as more words, and mirrored the child's utterances more compared to mothers of less verbal 2-year-olds. Similarly, Huttenlocher, Haight, Bryk, Seltzer, and Lyons (1991) showed that vocabulary growth between 14 and 22 months depends crucially on the quantity of parental speech input. Taken together, learning here is characterized as a dynamic interaction between what the child selectively attends to in, ideally, an optimally rich environment.

In the current study, we focus on children's early word and action learning. Infants as young as six months have been shown to discriminate familiar from unfamiliar actions (Hunnius & Bekkering, 2010) and to understand their first words (Bergelson & Swingley, 2013). But do they show improved learning of information from one domain over the other? Several studies have shown that children from the age of 30 months exhibit similar learning patterns when learning word- and action-object associations (Childers & Tomasello, 2002, 2003, Dysart et al., 2016, Riggs et al., 2015). However, in these studies, children were presented separately with the novel word-object and action-object associations. There are few studies that have compared word-object and action-object learning when information from both domains is presented in the same context. For example, Hahn and Gershkoff-Stowe (2010) presented 4- to 5-year-olds with novel word-object and action-object associations in the same context and found that they learned both words and actions for objects (although they found better performance for actions than for words).

Another study investigated the development of children's word- and action-object association learning by presenting 12-, 24- and 36-month-olds, as well as adults, with two novel objects which were presented with a novel label and a novel action. The following test phase measured their fixations to the target object upon hearing the label for this object or seeing the action that had previously been per-

formed on this object. 12-month-olds learned action-object but not word-object associations, 24-month-olds learned neither word-object nor action-object associations, 36-month-olds learned word-object associations, and adults learned both word-object and action-object associations (Eiteljoerge et al., 2019b). These results highlight the developmental dynamics of early word and action learning with differences in children's selective attention to different aspects of the signal potentially driving the differences in learning across development. The authors suggest that early in life children may prioritise attention to actions due to the increased salience of these movements, while later on, with greater language exposure, they may attend more to linguistic input due to the communicative salience of this domain. While these differences in selective attention were explained at the group level at different ages, the current study examines whether individual differences in selective attention to different aspects of the input may yield a similarly diverse pattern of learning from the two domains. To examine this, we required a paradigm that would allow children to demonstrate their current preference for words versus actions in the input available to them. Gaze-contingent paradigms may provide one possible solution to this problem.

In gaze-contingent paradigms, the participant can actively trigger a stimulus change with her own eye gaze: While the participant's eye gaze is tracked, live calculations in the background allow for an immediate analysis of the participant's looking behaviour, which can then be used to display stimuli in a reactive manner by allowing continued fixations at pre-specified areas on the screen to trigger subsequent stimulus presentation. Reder (1973) provides one of the earliest systematic descriptions of gaze contingency, showing how this methodology could be used for studies exploring the participant's active participation and control monitoring in a task.

More recent work shows how this paradigm can provide insights into children's early anticipation behaviour. For example, Wang et al. (2012) presented 6- and

8-month-olds with a dot on one side of the screen. When participants fixated the dot, an image would appear on the other side of the screen. Their study showed that as early as three trials into the paradigm, infants learned to actively control the stimulus presentation and anticipated the according outcome. Further, Deligianni et al. (2011) showed that 8-month-olds would exhibit gaze following to a contingent non-human object suggesting that the reactivity in a situation, whether human or non-human, can lead to infants' responsive behaviour. Thus, from early on, infants' gaze can be used to control the presentation of stimuli such that the infant is an active participant in shaping her learning experience.

Current study

In the current study, we investigated how children's (and adults') preferences for input from one or the other domain can influence their learning outcomes in either domain. In a gaze-contingent eyetracking paradigm, participants could choose between hearing the label for or seeing an action being performed on the same object. For example, a child would consistently be presented with the label for the object if she looked to the image of the object presented on the left side of the screen, and see an action for the object if she looked to the image of the same object presented on the right of the screen.

In a first phase, all objects presented were highly frequent objects likely to be familiar to the participant. This allowed us to establish children's preference for information from either domain across a range of objects presented on-screen. In a second phase, the participants were presented with images of novel objects, and could again choose the information they were provided about these objects by looking at the image of this object either on the left or the image of the same object on the right side of the screen. This allowed children to control what they learned about particular objects since the amount of exposure they received in either domain was entirely contingent on where they looked during the trial. In a following non-

gaze-contingent test phase, we examined children's learning and retention of the word- and action-object associations presented in the second gaze-contingent phase where they encountered the novel objects.

We hypothesised that children's learning of associations in both domains would depend on their preferences: a higher preference for information from one domain should result in better learning of this information in the novel association learning task. This would be in keeping with the literature on active learning described above suggesting that children follow their interests in selectively attending to their environment and retain information in keeping with these interests better. In contrast, we expected adults to learn information from both domains independent of any preferences they may have due to their potentially superior abilities to retain and process complex information and their success in a similar task in previous studies (Eiteljoerge et al., 2019b).

We also examined children's pupillary responses to violations of the associations they had been trained on, namely, violation of the actions performed on the different objects and violation of the kind of information presented on the different sides of the screen. We included the former test because 2- and 3-year-olds' inability to recognize trained action-object associations in the previous literature (Eiteljoerge et al., 2019b) could either suggest that children did not form a strong association between the action and the object during the training phase or that children did not retain and store this association strongly enough to access it during test in a task where another recently familiar competitor object was concurrently presented. To tease these two explanations apart, we presented children with trials where they either saw the same action being performed on an object that they had previously been trained on (match trials), or saw the different (but equally familiar) action being performed on this object (mismatch trials). Increased pupillary dilation to mismatch trials due to the violation of their expectation of the action performed on this object could then be taken to suggest that children learned and retained

the association but were unable to access this information in the context of the competitor object (Jackson & Sirois, 2009, Sirois & Jackson, 2012).

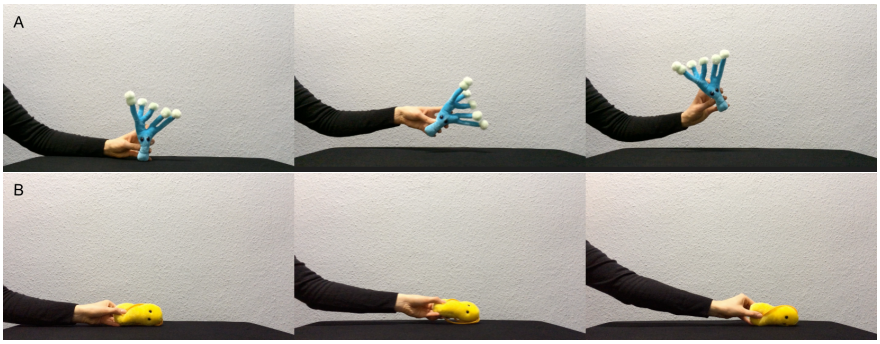
Second, we used an adapted version of this paradigm to test whether the participants associated information from one domain (e.g., words) with the side of the screen that information from this domain had been routinely presented on (e.g., left). Here participants saw two images of the same novel object on-screen. Fixations to a particular side of the screen triggered either the presentation of information from the domain that fixations to this side triggered in the earlier phases of the experiment (match trials), or the presentation of information from the other domain (mismatch trials). Increased pupillary dilation to mismatch trials would indicate that children had formed an association between the information domain and screen side during the previous trials and may have been sensitive to the fact that looks to different sides of the screen triggered different kinds of information.

Methods

Participants. Forty German monolingual 2- to 3-year-old children (range = 24;36 - 34;88 months; mean = 29;13, female = 22) and 26 adults (range = 19;34 - 32;07 years; mean = 24;19, female = 22) participated in the study. Additional participants were tested but excluded from the analysis during piloting (Children: 11), because of unwillingness to participate (Children: 4), technical failure (Children: 10, Adults: 1), preterm birth (Children: 3), experimenter error (Adults: 1) or insufficient data (Children: 9; Adults: 5, see Preprocessing). Children were recruited from the babylab database and participation was rewarded with a book. Adults were students of the University and were rewarded with 0.5 course credit points. Ethics approval was granted by the Ethics Committee of the Institute (Project 123).

Material. For the familiarisation phase, we used four familiar objects (a rabbit, a frog, a ball, and a plane), whose labels are typically familiar to children at the age of two ("Hase", "Frosch", "Ball", and "Flugzeug"), and typical actions which

Figure 5. Blue and yellow germ toys were used as novel objects. As novel actions, an upward movement with tilting to the sides (Panel A), and a sideways movement with tilting backwards and forwards (Panel B) were used.



can be performed on these objects (rabbit: hopping up and down, frog: hopping across the display, ball: bouncing, and plane: flying). We then created two types of videos: In the word videos, the label was presented in a carrier phrase (e.g., "Schau mal, ein Hase!" meaning "Look, a rabbit!") with the stationary object on screen. In the action videos, the object was presented in silence while a hand performed the respective action on it.

For the word learning phase, we used the same material as in Eiteljoerge et al. (2019b). It included two novel germs (`https://www.giantmicrobes.com/us/`, see Fig. 2), two novel labels that adhere to German phonotactic rules ("Tanu" and "Löki"), and two novel actions ("moving up and down while tilting to the sides" and "moving sideways with back and forward leaning"). We again created two types of videos: In the word videos, the object was displayed as a still frame while the label was presented in a carrier phrase (e.g., "Schau mal, ein Tanu!" meaning "Look, a Tanu!"). In the action videos, the object was presented in silence while moving according to the assigned action. All videos were roughly seven seconds long, and both words and actions were only presented once per video to match the exposure across domains (materials can be found on OSF under `https://osf.io/xeqb2/`).

For the test phase, we used pictures of the two objects in the baseline and the test phase. These pictures were 640×480 and presented in the center of the left and the right half of the screen. Thus, in a first baseline phase, the two pictures were

presented on screen in silence, Then, in a prime phase, either a recording directed children to look at the one of the objects using its label “Where is the Tanu?” or a video presented a hand performing the action associated with one of the objects (without the object) was played. These prime videos were presented in the center of the screen and were roughly seven seconds long. In the test phase, both objects were presented on screen again in the same positions as before.

For the action violation phase, we used videos of the action presentation without any linguistic information. These videos were seven seconds long, and presented in the center of the screen. For the side violation phase, we used the same material and the same display as in the familiarisation phase. For the pupil dilation measure during both the action and the side violation phase, we used a 10 s video of bubbles that was stable in luminosity and kept the participants’ attention on screen, courtesy of Hepach, Vaish, and Tomasello (2012).

Procedure. Participants sat in a dimly lit and quiet experimental room at a distance of 65 cm from a TV screen (92 × 50 cm), where the visual stimuli were presented. A remote eye tracker (Tobii X 120, Manual, 2008), set on a platform underneath the TV screen, was used to record gaze data at 60 Hz. Stimuli were presented using E-Prime software with the Tobii extension (Schneider et al., 2002). Auditory stimuli were presented through two loudspeakers above the screen. Further, two video cameras centered above the screen served to monitor the participant during the experiment. Prior to testing, gaze was calibrated using a 5-point calibration procedure, in which a red point appeared in every corner and the center of the screen. The calibration was checked using two pictures on the screen (a snail and a frog). Any systematic divergences in looking to the two pictures would lead to recalibration. After successful calibration, the experiment started with a fixation getter in the middle of the screen before each stimulus presentation.

Experimental design. All participants went through five different phases. The total length of the study depended on the participants (due to the gaze-contingent design) but usually took around ten minutes.

Gaze-contingent preference phase. The first eight trials of the experiment presented the participant with a still image of the same picture of a familiar object on the left and on the right side of the screen. When the participant looked at one of the two images, a frame in light gray would appear around this picture. If the participant then looked to the other side, this frame would follow to the other side. If the participant fixated one of the two images for 1000 ms, the other image would disappear and the chosen video would start to play. Importantly, one side was always associated with the same information content for individual participants (e.g., word video on the left side, action video on the right side). Thus, by fixating one side for at least 1000 ms, the participant could trigger the presentation of a video in the preferred domain. Eye blinks or short shifts initiated an additional counter. If the time period of looking away did not exceed 160 ms, counting towards the current fixation was continued. If it exceeded this time window, the counter for a fixation was reset. If the participant did not fixate any of the pictures for 15 s, the trial was aborted. This phase was meant to familiarise the participant with the gaze-contingent procedure, establish the side association of word and action presentations, and measure the preference of the participant for either word or action presentations.

Gaze-contingent exposure phase. The gaze-contingent exposure phase presented the participant with the possibility to learn novel words and novel actions for novel objects. The presentation of the stimuli followed the same paradigm as in the preference phase. This time however, the participant was presented with still images of the novel objects. A fixation of 1000 ms to either the word or the action side would trigger the presentation of their novel labels or the corresponding actions performed on them, respectively. Children were presented with a total of 16 trials

(eight per object), while adults were presented with a total of four trials (two per object).

Test phase. In eight following test trials, the two novel objects were presented side by side on screen in silence for 2.5 s (baseline phase). The objects disappeared after 2.5 s, triggering the beginning of a prime phase (similar to Holzen & Mani, 2012), during which participants either heard the label for one of the objects (Word-Object trial) or saw an empty hand performing one of the actions (Action-Object trial). 300 ms after the offset of the label or the end of the action, the objects reappeared on-screen for a further 2.5 s in the same positions as in the baseline phase (test phase). This test phase was identical to the test phase in the previous study on word and action learning (Eiteljoerge et al., 2019b).

Action violation phase. In four further trials, we examined alternative indices of participants' action-object learning. Here, we presented participants with two match trials, in which the participants saw the novel objects and the previously associated actions (e.g., blue object moving up and down as in the exposure phase), followed by two mismatch trials in which the two objects were presented with the action (e.g., blue object moving to the side). Furthermore, a video of bubbles was presented before and after each trial, during which we measured tonic changes in pupil dilation (Hepach et al., 2012).

Gaze-contingent side violation phase. With the last four trials, we wanted to examine whether participants associated a particular type of information with a particular side of the screen during the familiarisation phase. In other words, we wanted to examine here whether those participants who were always provided with word information on the left and action information on the right were sensitive to this assignment of side and information content. Such sensitivity would be key to any conclusions regarding participants' active manipulation of their learning experience. Thus, we presented participants with two side-match trials, where the word and action information were presented on the same side that they had been pre-

sented on during the familiarization phase, and two side-mismatch trials, where the word and action information were presented on the opposite side that they had been presented on during the familiarization phase. As above, a video of bubbles was presented before and after each trial for tonic measures of pupil dilation following side-match and side-mismatch trials.

Pre-processing. The eye-tracker provides an estimate of where participants were looking at during each time stamp of the trial, with one data point approximately every 16 ms. All data (gaze data and trial information) were exported from E-Prime and then further processed in R (R Core Team, 2016). For each time stamp, data were only included when one or both eyes of the participant were tracked reliably (validity less than 2 on E-Prime scale). When both eyes were tracked, the mean gaze point for both eyes was computed for further analysis. Gaze data were then aggregated into 40 ms bins. Areas of interest identical in size to one another were defined based on the size and location of the videos and images on screen. For the test phase, we coded whether the participant looked at the correct object (i.e., the target), the distractor or at neither of these onscreen. Typically, fixation data is analysed from 240 ms to 2000 ms from the onset of a label (see also Swingley et al., 1999), since only these fixations can be reliably construed as a response to an auditory stimulus. Following this convention, we analysed the data from 240 ms from the onset of the baseline phase. This ensured that we avoided early fluctuations in fixations as a result of the images appearing on-screen. Furthermore, we corrected the test phase for any preferences in the baseline phase on an individual trial level (i.e., each time point in the test phase was corrected for the overall looking score in the baseline phase on that particular trial).

For the pupil dilation measure, only timestamps providing valid eyetracking data were included in the analysis. We used a threshold filter for both pupils separately, that excluded data points in which the difference between two adjacent data points fell into the top ten percent of overall differences between adjacent time

points. This procedure ensures that large deviations between adjacent samples are excluded which are often artefacts in the data. Further, data were interpolated with a sample size of four to account for gaps due to missing data (Hepach et al., 2012). Finally, data from both eyes were averaged.

To ensure good data quality, we used specific exclusion criteria. Participants who did not trigger at least one action or word video in the entire gaze-contingent preference phase were excluded from the study as we could not be sure that these participants had understood the study design (C: n = 8, A: n = 3), since they had never seen the other information that could be provided. Further, one participant who did not trigger any content in any of the videos was excluded (A: n = 1). Also, single test trials were excluded if a participant was looking at the stimuli more or less than M ± 3SD of the time. Thus, a trial was excluded if a child looked less than 52.13 % of the time during the trial to one of the two pictures. This led to an exclusion of 20 trials (7.69 %). For adults, this criterion was at ≤ 38.93 %, 15 trials (7.58 %). This led to another exclusion as the participant did not look enough during any of the trials (A: N = 1). Furthermore, participants were excluded if they contributed trials in only one of the two conditions (C: N = 1). Finally, it was ensured that participants contributed more than one trial per condition in the test phase and that there were no missing data in either the Baseline or the Test phase in one of the conditions. This did not lead to further exclusions, and thus, left us with 66 participants overall.

Results

First, we present the analysis of target fixations during the test trial to examine whether participants had learned the word-object and action-object associations presented, and how their learning interacted with their choices in the preference phase. Then, we present the analysis of the action violation and the side violation phase. Finally, we present an exploratory analysis testing whether participants con-

tinued their preference during the familiarisation phase during the learning phase, or whether they changed their information sampling method when presented with new information in the learning phase.

The influence of preference on learning. In this section, we first report analyses of all age groups together, then split the data according to age, and by condition to test whether each age group learned in both of the conditions and whether there are differences between conditions. For each analysis, we report two statistical approaches. First, we report traditional t-tests and correlations. Then, we report generalized linear mixed models controlling for additional factors such as object, name, action, and exposure, which also allows us to include the factor *Time* during the trial. Our aim here is not to compare the characteristics of the individual types of analyses, but rather to ensure that the findings we report converge when using both statistical approaches.

We used generalised linear mixed models (GLMM) to analyse how participants' proportional target looking unfolded over the entire time course of the test phase (also called growth curve models; for a more detailed description and instruction see Mirman, 2014, Mirman et al., 2008). Were participants to not recognize the target, they ought to fixate both the target and distractor equally throughout the time course of the trial, with their fixations at chance altogether. Were participants to recognize the target, their fixations to the target should increase from chance level during the trial, and return towards chance level towards the end of the trial (mirroring a quadratic curve, or a quartic curve when considering tails). In practice, adults usually fixate the target until the end of the trial (e.g., Mirman et al., 2008), while children's target looking appears to increase later during the trial, and is often weaker and more variable throughout the trial (see Eiteljoerge et al., 2019b, for a more elaborate description).

We fitted a generalized mixed model using lme4's lmer function in R (Bates et al., 2015) with Gaussian error structure and identity link function. For all models,

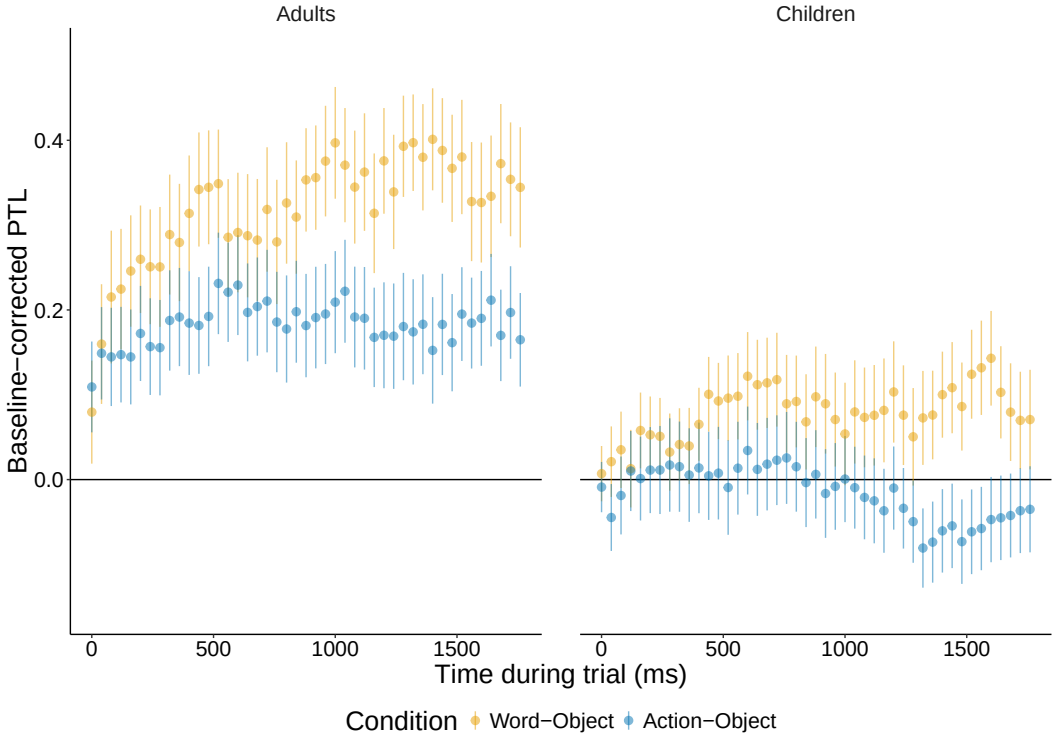
participant id, label (Tanu, Löki), object colour (blue object, yellow object), and action (up, side) were included as random effects. Age at test, preference, and exposure were z-transformed (to a mean of zero and a standard deviation of one) to simplify later interpretation across groups. The exact models for each analysis can be seen below.

We highlight here a difference in our calculations from the gaze-contingent preference and the gaze-contingent exposure phase. We refer to the choices made my participants in the preference phase as their preference for actions versus word information associated with objects in general. We refer to the choices made by participants during the exposure phase as exposure, since these choices determined the number of word and action trials participants saw, training them on the novel word-object and action-object associations, respectively. While it is likely that preference and exposure thus defined are related to one another, we use this notation to differentiate between the two phases and to motivate some of the analyses that follow. For each phase and each domain, we calculated the overall number of trials chosen in that domain.

We also included *Time* and its second, and third polynomial in the model. This allowed us to model our data as a linear, quadratic, and cubic function of time, and thus, to capture according shapes in the data (the number of polynomials has been restricted to four based on theoretical considerations regarding looking behaviour). We used the function *drop1* to evaluate the influence of each factor in the model. This function compares the model including one factor with a model without this factor, and thereby evaluates its contribution to the model. An overview of the results can be seen in Tab. 9, Tab. 10, Tab. 11, and Fig. 6, Fig. 7, and Fig. 8.

Both age groups. We first run a linear mixed model to evaluate influences of condition and age on target looking. The model revealed a main effect of Condition, $\chi^2 = 8.26$, df $= 1$, $p = .004$, and a main effect of Age, $\chi^2 = 13.32$, df $= 1$, $p < .001$.

Figure 6. Time course graphs for participants' baseline-corrected proportional target looking (PTL) in the Word-Object and the Action-Object condition during the test phase after label onset and 233 ms to initiate a fixation. The line at 0 represents chance level, looks above the line reflect proportionally more target looking and looks below the line reflect distractor looking.



These results suggest differences across age groups and across conditions (see Fig. 6. We then split the data according to age group.

2- to 3-year-olds. We ran a linear mixed model to test the effect of condition on target looking. The model revealed a main effect of Condition, $\chi^2 = 4.4$, df = 1, $p = .036$, with higher proportional target looking in the word compared to the action condition.

Then, we split the data according to condition and added the preference and exposure for the respective condition to the model. Thus, for the word-only model, we analysed the influence of word preferences and word exposure on target looking, while for the action-only model, we analysed the influence of action preferences and action exposure on target looking.

For the word-only model, the model revealed a significant effect on the intercept (p = .005), indicating that children's target looking in the word-object condition differed from chance. For the action-only model, none of the factors were significant, indicating that children's target looking in the action-object condition did not differ from chance.

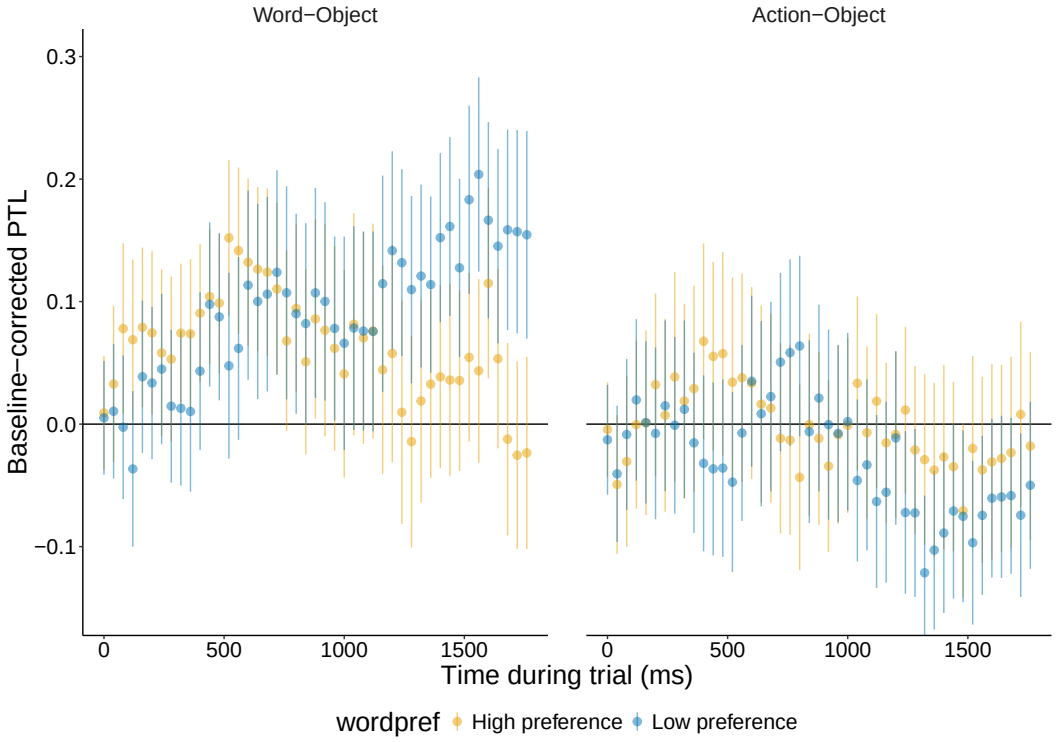
These results suggest that 2- to 3-year-old children learned to associate words with objects but not actions with objects, and word and action preferences did not influence their target recognition in the two domains.

To evaluate how participants' choices in the preference phase influenced the time course of participants' target looking, we run exactly the same models as the previous ones that included preferences and exposure separately for each condition, and additionally included polynomials up to the cubic term and their interaction with the preference score.

Using drop1, the word-only model revealed a significant interaction of word preference*poly3 (χ^2 = 6, df = 1, p = .014). The action-only model revealed no significant interactions between preference and polys, indicating that children's time course of target looking in the action-object condition did change with their level of action preference.

Thus, on the group level, 2- to 3-year-old children showed a difference between word and action learning, with proportionally more target looks in the word-object than in the action-object condition. Furthermore, they learned to associate words but not actions with objects. On an individual level, preferences influenced target looking in the word-object but not in the action-object condition. Children with a high word preference showed a cubic curve that increased from chance level and returned to chance level towards the end of the trial. Children with a low word preference showed a steady increase in target looking throughout the trial (see Fig. 7).

Figure 7. Time course graphs for 2- to 3-year-olds' baseline-corrected proportional target looking (PTL) in the Word-Object condition during the test phase after label onset and 233 ms to initiate a fixation. For visualisation purposes, data were split according to the median of word preference. A high word preference is represented in yellow, a low word preference is in blue. The line at 0 represents chance level, looks above the line reflect proportionally more target looking and looks below the line reflect distractor looking.



Adults. We ran a linear mixed model to test the effect of condition on target looking. The model revealed a main effect of Condition, $\chi^2 = 4.11$, df $= 1$, $p = .043$, with higher proportional target looking in the word than in the action condition.

Then, we split the data according to condition and added the preference and exposure for the respective condition to the model. Thus, for the word-only model, we analysed the influence of word preferences and word exposure on target looking, while for the action-only model, we analysed the influence of action preferences and action exposure on target looking.

For the word-only model, the model revealed a significant effect on the intercept ($p < .001$), indicating that target looking in the word-object condition differed from chance. Also for the action-only model, the model revealed a significant effect on the

intercept (p = .001), indicating that target looking in the action-object condition differed from chance.

These results suggest that adults learned to associate words and actions with objects, and word and action preferences did not influence their overall target looking in the two domains.

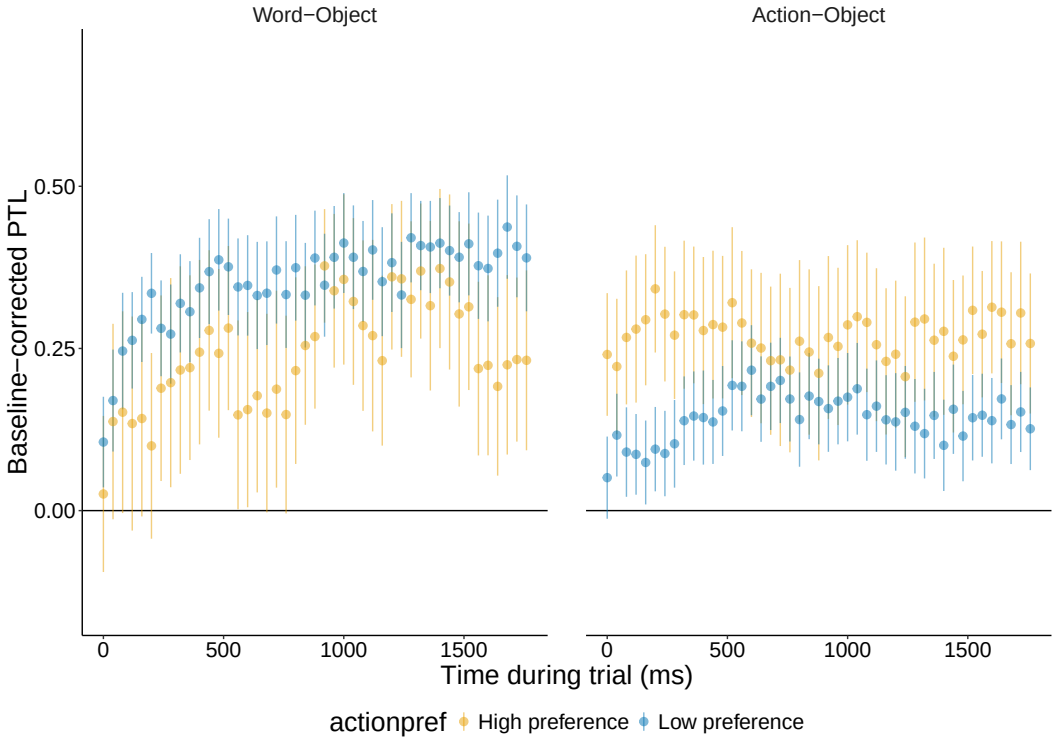
To evaluate how participants' choices in the preference phase influenced the time course of participants' target looking, we run exactly the same models as the previous ones that included preferences and exposure separately for each condition, and additionally included polynomials up to the cubic term and their interaction with the preference score.

Using drop1, the word-only model revealed no significant interactions between word preference and the polys, indicating that the time course of target looking did not change significantly with their preference. The action-only model revealed a significant interaction of action preference*poly3 (χ^2 = 4.46, df = 1, p = .035), indicating that the time course of target looking changed with their action preference.

Thus, on the group level, adults showed a difference between word and action learning, with proportionally more target looks in the word-object than in the action-object condition. Furthermore, they learned to associate words and actions with objects. On an individual level, preferences influenced target looking in the action-object but not in the word-object condition. Adults with a high action preference showed a high plateau of target looking throughout the trial. Adults with a low action preference showed a cubic curve of target looking: Their time course of target looking showed a slower increase and a lower plateau compared to adults with a high action preference (see Fig. 8).

Action violation. In the action violation phase, we tested whether participants were sensitive to a change in the action that was performed on the object. As a measure, we used participants' pupil dilation during the bubbles videos (Hepach

Figure 8. Time course graphs for adults' baseline-corrected proportional target looking (PTL) in the Action-Object condition during the test phase after label onset and 233 ms to initiate a fixation. For visualisation purposes, data were split according to the median of action preference. A high action preference is represented in yellow, a low action preference is in blue. The line at 0 represents chance level, looks above the line reflect proportionally more target looking and looks below the line reflect distractor looking.



et al., 2012) which were presented between the stimuli videos. As a baseline, we used the first bubbles video that was presented before any of the action videos. For the match trials, we then used the second and third presentation of bubbles that each followed after one of the videos that presented a matching action. For the mismatch trials, we then used the fourth and fifth presentation of bubbles that each followed after one of the videos that presented a mismatching action.

First, we report t-tests comparing match trials vs. baseline and mismatch trials vs. baseline. Then, we report linear mixed models that evaluate the change of pupil dilation in mismatch trials against chance level (chance = 0), while controlling for participants' pupil dilation in the match trials. This was done to incorporate any carry-over effects from the match trials due to the fixed order of match-mismatch

trials. To account for basic behaviours of the pupil in later trials, we baseline-corrected both match (match - baseline) and mismatch trials (mismatch - baseline). With regard to the models, an effect on the intercept would suggest a difference between the averaged minimum pupil dilation in mismatch trials and chance, while an effect of match trials would indicate an additional influence of match trials. Note that χ^2 statistics are not available for intercept statistics, because χ^2 is a measure based on model comparison. A model without an intercept is impossible, and therefore, there is no model comparison available.

Before the analysis, we ensured that each participant contributed data in the first and second window of baseline, match, and mismatch. Thus, for the following analyses, we could include 28 children and 22 adults.

Table 4
Minima of pupil dilation in the action violation phase for 2- to 3-year-olds and adults.

Age	Baseline	Match	Mismatch
Children	2.11	2.11	2.25
Adults	3.10	2.95	2.55

2- to 3-year-olds. For the 2- to 3-year-olds, a paired t-test between mismatch and baseline trials was significant (t(27) = -2.28, p = .031, d = -0.61), indicating an increase in pupil dilation in the mismatch trials compared to baseline. The comparison between match and baseline trials was not significant (t(27) = -1.44, p = .160, d = -0.39).

The model comparing pupil dilation in mismatch trials to chance while controlling for pupil dilation in match trials revealed a significant effect of pupil dilation in the mismatch trials on the intercept (p = .01) and of match trials (χ^2 = 10.53, df = 1, p = .001). Pupil dilation was significantly above chance in mismatch trials, and larger pupil size in match trials was related to larger pupil size in mismatch trials.

The increase in pupil dilation from chance level in mismatch trials suggests that 2- to 3-year-old children were sensitive to the action violation.

Adults. For the adults, a paired t-test between match and baseline trials was significant (t(21) = 3.41, p = .003, d = 1.03), as was the comparison between mismatch and baseline trials (t(21) = 2.35, p = .028, d = 0.71), indicating a decrease in pupil dilation in both match and mismatch trials compared to baseline.

The model comparing pupil dilation in mismatch trials to chance while controlling for pupil dilation in match trials revealed a significant effect of pupil dilation in the mismatch trials on the intercept (p = .023), and a significant effect of match trials (χ^2 = 16.8, df = 1, p < .001). Pupil dilation was significantly below chance in mismatch trials, and larger pupil size in match trials was related to larger pupil size in mismatch trials.

The decrease in pupil dilation from chance level in mismatch trials suggests that adults might have shown symptoms of disengagement and fatigue during the task (Hopstaken, van der Linden, Bakker, & Kompier, 2015).

Side violation. In the side violation phase, we tested whether participants associated one type of content (i.e., words) with one side of the screen (i.e., left) during the preference phase, and would therefore be sensitive to a violation in the presentation (i.e., words presented on the right). Again, we used the bubbles video and participants' change in pupil dilation to test this aspect, and we report the same analyses as in the section above.

Before the analysis, we ensured that each participant contributed data in the first and second window of baseline, match, and mismatch. Thus, for the following analyses, we could include 32 children and 25 adults.

Table 5
Minima of pupil dilation in the side violation phase for 2- to 3-year-olds and adults.

Age	Baseline	Match	Mismatch
Children	2.44	1.97	2.48
Adults	2.96	2.99	2.07

2- to 3-year-olds. For the 2- to 3-year-olds, a paired t-test between match and baseline trials was significant (t(31) = 2.17, p = .038, d = 0.54), indicating a decrease in pupil dilation from baseline to match trials. Further, a paired t-test between match and baseline trials was significant (t(31) = 3.24, p = .003, d = 0.81), indicating an increase in pupil dilation from baseline to mismatch trials.

The model comparing pupil dilation in mismatch trials to chance while controlling for pupil dilation in match trials only revealed a significant effect of match trials (χ^2 = 22, df = 1, p < .001). Pupil dilation was not significantly different from chance in mismatch trials, and larger pupil size in match trials was related to larger pupil size in mismatch trials.

These results suggest that children were not sensitive to the side violation.

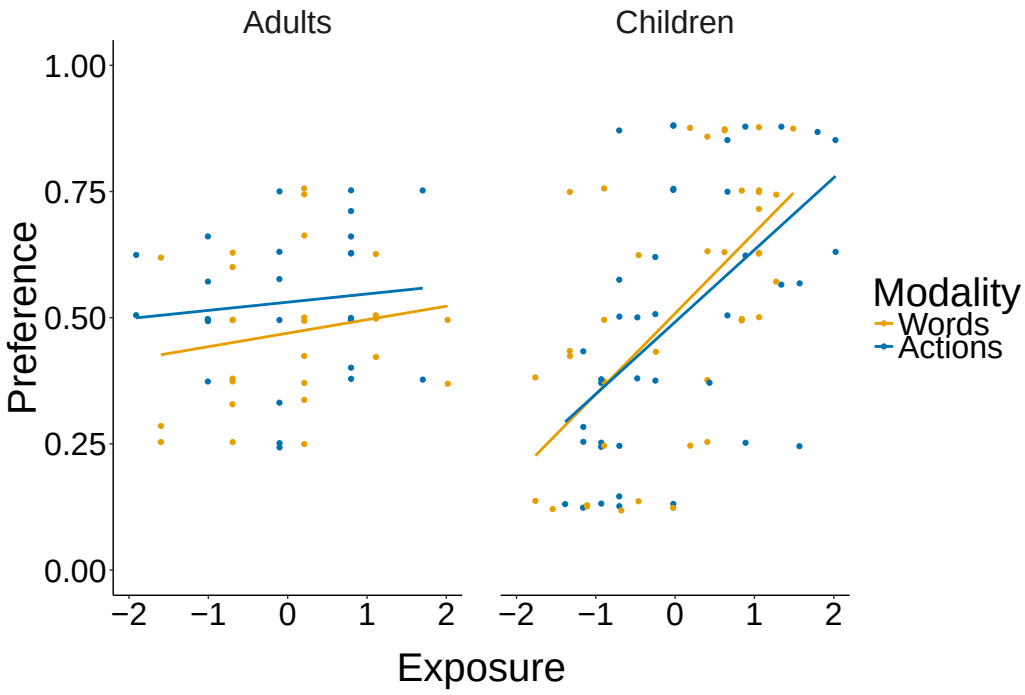
Adults. For the adults, a paired t-test between match and baseline trials was not significant (t(24) = 1.86, p = .076, d = 0.52), and neither was the comparison between match and baseline trials (t(24) = 0.76, p = .45, d = 0.22), indicating that the pupil size did not change significantly across trial types.

The model comparing pupil dilation in mismatch trials to chance while controlling for pupil dilation in match trials revealed only a significant effect of match trials (χ = 8.02, df = 1, p = .005). Pupil dilation was not significantly different from chance in mismatch trials, and larger pupil size in match trials was related to larger pupil size in mismatch trials.

These results suggest that adults were not sensitive to the side violation.

Further exploratory analysis. To investigate whether preference and exposure choices aligned for children and adults, we correlated these two measures (Fig. 9). For the 2- to 3-year-olds, preference and exposure choices were positively correlated for words (r(40) = .61, p < .001) and for actions (r(40) = -.55, p < .001), meaning that higher preference values in one domain meant higher exposure values in the same domain. For adults, both correlations were not significant.

Figure 9. Correlation of word and action choices in the preference and the exposure phase for 2- to 3-year-olds and adults. Note that the exposure score has been z-transformed to allow for comparisons across groups.



Discussion

In the current study, we investigated the influence of preference on learning of word-object and action-object associations in 2- to 3-year-old children and adults using a gaze-contingent paradigm. In a first gaze-contingent preference phase, participant could choose with their gaze whether they wanted to hear a familiar label or see a familiar action for a familiar object. In the following gaze-contingent exposure phase, participants could again choose whether they wanted to hear a novel label or see a novel action for a novel object. Afterwards, their learning of the presented word-object and action-object associations was tested in a preferential looking paradigm. We found that children at this age learned word-object but not action-object associations, replicating our previous findings (Eiteljoerge et al., 2019b). Further, children's preferences influenced target looking in the word-object

condition but not in the action-object condition. Adults' preferences influenced target looking in the action-object condition but not in the word-object condition.

Our results on children's word and action learning are in line with our previous study in which we found that 3-year-olds would learn word-object but not action-object associations (Eiteljoerge et al., 2019b). We extended these results showing how children's individual preferences can modulate their behaviour, and found that only their target looking in the word-object condition, but not in the action-object condition, was modulated by their preferences. Children with a high word preference showed an increase in target looking from chance level in the beginning of the trial, which decreased towards the end of the trial, whereas children with a low word preference showed a steady increase in target looking throughout the trial. Thus, the pattern of target looking in the children with a high word preference resembles a pattern that would be expected on a theoretical level (as outlined above), and might therefore be taken as evidence for children's successful target recognition in word-object test trials. For children with a low word preference, their target looking appears to increase later. Nonetheless, it is important to note that, overall, children between two and three years of age learned words for objects successfully. These observations were mirrored in the action-object condition for the adults: They learned action-object associations, but their target looking at test followed different trajectories with differences in preference. Adults with a high action preference showed a high plateau throughout the trial, while adults with a low action preference showed a slower increase with a lower plateau. Both patterns are indicative of adults' action-object association learning, and expected given adults' typical target looking behaviour (Mirman et al., 2008). Together, these findings can be taken as support for the conclusions by Wade and Kidd (2019) who suggest that curiosity may contribute to learning, but its contribution is limited: In our setting, preferences appeared to change the response behaviour but did not change the overall learning pattern.

The influences of preference on target looking differed from our expectations in the two age groups. While we expected an influence in both domains for children, we expected no influence for adults. One possible explanation for the diverging results might be the relation between curiosity and participants' certainty in knowing an association. Adults' target looking in the word-object condition was overall high and appeared to have little variance, adults' target looking in the action-object condition and children's target looking in the word-object condition appeared to have high variance, whereas children's target looking in the action-object condition was overall around chance level and appeared to have little variance. If we take variance as an indicator of certainty, then increased variance in the data might be a sign for higher uncertainty. Accordingly, in scenarios with high variability, preferences might have stronger influences on behaviour compared to scenarios where uncertainty is low, because such preferences might help to reduce uncertainty during information processing (Gottlieb, Oudeyer, Lopes, & Baranes, 2013).

The influence of preference on participants' target recognition reflects how characteristics of the learner contribute to learning: In terms of dynamic systems accounts, this reflects the interactive nature of the organism and the environment (e.g., Smith & Thelen, 1993). While the environment provides information (e.g., in form of words and actions for objects), the learner's preferences play an active role in the processing of information. In this way, the learner can focus on specific aspects of the input, which in turn helps them to reduce the information to be processed (Kidd & Hayden, 2015).

In contrast to children's word-object association learning results, we did not find evidence for children's learning of action-object associations which is in line with our previous results (Eiteljoerge et al., 2019b). In the current setting, action-object learning might have been impeded additionally by the interleaved presentation of these associations that resulted from children's choices in the exposure phase. This could result in large gaps between the same input, and thereby impede learning of

a certain mapping at this age. Accordingly, blocking the presentation of the associations might simplify learning of the associations in the current setting (Mather & Plunkett, 2011). Older children who have been shown to learn both words and actions for objects (e.g., Hahn & Gershkoff-Stowe, 2010) might show action-object association learning because of their further developed perceptual abilities. The absence of an influence of preference in the action-object condition could be due to the smaller amount of variance in this condition as described above: if most children did not show action-object learning at all, any preferences might have little effects on target looking modulations. Thus, if older children learn the action-object associations, any potential influences of preferences on learning might then surface.

Although children did not show target recognition in the action-object condition at test, they seemed to discriminate between match and mismatch trials in the action violation phase, as indicated by an increase in pupil dilation. This suggests that they were able to recognize a change in the action-object association despite not storing this information for longer. This is in contrast to adults, who showed successful learning of action-object associations in the test phase, but showed a decrease in pupil dilation as a reaction to the action violation. These results potentially shed light on the different nature of recognition and learning (McMurray, 2016). For children, recognising a change in action-object associations might involve low processing demands while recalling them in the test phase later in the experiment might have required more efforts to store the mapping for longer. For adults however, learning an action-object association might be less demanding. In addition, adults might be more tolerant regarding different actions for the same object based on their experience: in contrast to many word mappings, action mappings are often not necessarily one-to-one mappings. Accordingly, they might rather show symptoms of disengagement and fatigue (Hopstaken et al., 2015) instead of an increase in pupil dilation which could reflect the violation of an expectation (Satterthwaite et al., 2007).

For our gaze-contingent paradigm to work, it was important that participants recognized that each side of the screen was associated with one type of information. However, both age groups did not show any changes in pupil dilation after the mismatch trials. These results could indicate that participants did not associate one type of content with one side of the screen. The positive correlation between the preference and the exposure phase suggest a systematic pattern in children's choices which could be taken as indirect evidence for their recognition of the side association. However, these results could also reflect a side preference that was present in children but not adults. Although we coined participants' behaviour as *choice* indicating a *preference* throughout the manuscript, these results might also support a leaner interpretation of their behaviour in terms of selective attention to the information presented or a side preference in the case of children. We asked some adults after the experiment whether they noticed something, and they often mentioned that they figured out that they could choose with their gaze which could be taken as anecdotal evidence for adults' awareness of their choice during the preference and exposure phase. However, participants' awareness of the functioning of the paradigm and the underlying mechanisms during such gaze-contingent study designs still require substantial investigation. If we can confirm in future studies that participants are aware of their choices and indeed choose according to their preference, such gaze-contingent designs could be an interesting avenue for research that tries to bridge highly controlled experimental settings and purely observational tasks, allowing for many more research questions that address the interactions between the learner and the environment.

To conclude, we found that 2- to 3-year-old children learned words but not actions for objects, and that their preferences influenced target looking in the word-object but not in the action-object condition. But although they did not show action-object association learning at test, they were sensitive to a change in action-object mappings, suggesting that they remembered action-object mappings from

one trial to the next. Further, we found that adults learned words and actions for objects, and their preferences influenced target looking in the action-object but not in the word-object condition. Overall, this research contributes to the growing literature on how children's own characteristics like their preferences can influence their behaviour, and how this influence changes with development.

Consistency of co-occurring actions influences young children's word learning

Abstract

Communication with young children is often multi-modal in nature, involving, for example, language and actions. The simultaneous presentation of information from both domains may boost language learning by highlighting the connection between an object and a word, owing to temporal overlap in the presentation of multimodal input. However, the overlap is not merely temporal but can also co-vary in the extent to which particular actions co-occur with particular words and objects, e.g., caregivers typically produce a hopping action when talking about rabbits and a snapping action for crocodiles. The frequency with which actions and words co-occur in the presence of the referents of these words may also impact young children's word learning. We therefore examined the extent to which consistency in the co-occurrence of particular actions and words impacted children's learning of novel word-object associations.

Children (18 months, 30 months, 36 - 48 months) and adults were presented with two novel objects and heard their novel labels while different actions were performed on these objects, such that the particular actions and word-object pairings always co-occurred (*Consistent* group) or varied across trials (*Inconsistent* group). At test, participants saw both objects and heard one of the labels to examine whether participants recognised the target object upon hearing its label.

Growth curve models revealed that 18-month-olds did not learn words for objects in either condition, and 30-month-old and 36- to 48-month-old children learned words for objects only in the *Consistent* condition, in contrast to adults who learned words for objects independent of the actions presented. Thus, consistency in the multimodal input influenced word learning in early childhood but not in adulthood. In terms of a dynamic systems account of word learning, our study shows how

multimodal learning settings interact with the child's perceptual abilities to shape the learning experience.

Introduction

Infants grow up in a multimodal world, where language input is typically embedded in a contextual framework. For example, this input can be provided with concurrently presented gestures like pointing and actions like the hopping of a rabbit. Indeed, words are often accompanied by actions (Matatyaho-Bullaro et al., 2014), and actions are often accompanied by words (e.g., Elsner & Pfeifer, 2012, Southgate, Chevallier, & Csibra, 2009). In a study by Gogate and colleagues, temporal alignment of parental language and actions in naming events was around 76% for preverbal infants, underscoring the co-occurrence of speech and action in early communication with infants (Gogate et al., 2000). This rich environment provides infants with the opportunity to experience and learn from different sources and senses but also requires the learner to play an active part in shaping her learning experience, filtering out what is relevant for her. How does a child determine what to focus on when confronted with simultaneous speech and action? Does information from one domain interfere with the processing in the other domain, or does cross-domain information foster processing on another domain? We examined these questions in the current study by investigating the extent to which consistency in the co-occurrence of particular actions and word-object pairings influenced children's learning of the novel word-object associations: in a word learning task, two groups of 18-, 30-month-old and 3- to 4-year-old children and adults saw two objects and heard their labels while the objects moved in either the same (*consistent* group) or in a varied (*inconsistent* group) manner across successive presentations of the word-object association. At test, participants' learning of the word-object associations was tested in a preferential looking paradigm (Hirsh-Pasek & Golinkoff, 1996) to see whether action consistency had an influence on their word learning.

The literature on early speech perception and word learning documents the impressive pace and flexibility with which infants acquire their native language, typically understanding and producing their first words even before their first birthday (e.g., Benedict, 1979, Bergelson & Swingley, 2012). Similarly, infants display rapid development in the action domain from early on. By 3 months, infants can represent goal-based actions (Sommerville, Woodward, & Needham, 2005), and at least by 9 months they seem to understand the implications of actions (Behne, Carpenter, Call, & Tomasello, 2005). Thus, already at 1 year of age, infants show impressive language and action processing skills.

Further, words and actions often co-occur in the input: as mentioned above, Gogate and colleagues found in their study that temporal alignment of actions accompanying caregiver speech was around 76% during year 1 (Gogate et al., 2000), and 2-year-old children accompany their own speech with actions (Nicoladis, Mayberry, & Genesee, 1999). Much work by Gogate and colleagues has investigated the role of mothers' action presentations during semi-structured free play and finds that mothers favor certain action types in temporal synchrony with language in communicative interactions with their 6- to 8-month-olds (Matatyaho & Gogate, 2008). With time, the temporal synchrony of actions and words in maternal communication with infants reduces, adapting to the child's developmental needs and their reduced reliance on synchronicity in multimodal interactions (Gogate et al., 2000).

The high co-occurrence of words and actions in the input of the child might be accompanied by cross-domain influences on word and action learning. Indeed, studies have shown that young infants seem to benefit from multimodal input when learning novel words. In these studies, the concurrent presentation of actions and words or syllables helped 6- to 8-month-olds to learn the word-object mappings, as long as actions and words were temporally synchronous and infants attended to the presentation of the multimodal input (e.g., Gogate, Bolzani, & Betancourt, 2006).

Similarly, in experimentally controlled habituation paradigms, 8-month-olds were better able to learn word-object mappings when the objects were presented with temporally synchronous actions, with some actions being favored compared to others (e.g., *shaking* compared to *sideways,* Matatyaho-Bullaro et al., 2014). Indeed, Werker et al. (1998) report that 14-month-olds learned the associations between words and their referents *only* when these objects were moving, but not when they were stationary. These results suggest that concurrently presented actions foster word learning in young children, potentially due to such multimodal content highlighting the connection between the word and the object. In contrast, Puccini and Liszkowski (Puccini & Liszkowski, 2012) find that 15-month-olds learned words for objects, but not when they were accompanied by a referential gesture (e.g., a fist moving up in down). Like the study by Matatyaho-Bullaro et al. (Matatyaho-Bullaro et al., 2014), these results suggest that the nature of the concurrently presented actions or, in this case, gestures, may influence young children's word learning.

Likewise, language input can influence children's interpretation of action demonstrations. For example, language can help structure action sequences (i.e., acoustic packaging, Brand & Tapscott, 2007, Meyer, Hard, Brand, McGarvey, & Baldwin, 2011) and modulate children's representation of goal-directed actions (Elsner & Pfeifer, 2012) by highlighting the relevance of an action and guiding children's imitation of these actions (see also Southgate et al., 2009). Language can also facilitate the comparison of actions in infants early as 10 months of age and help infants to understand actions as being goal-directed (Gerson & Woodward, 2014).

Thus, the concurrent presentation of information from the word and action domains fosters learning in both domains. The studies reviewed above have all, however, examined the temporal synchrony between (particular) action demonstrations and language input, i.e., the effect of actions being presented concurrently with language. Here, we extend this to examine the influence of the consistency of

concurrently presented action demonstrations on word learning. By consistency we refer to the consistency with which certain actions accompany certain word-object associations. For example, when parents introduce children to new word-object associations, parents might make the rabbit hop or wiggle its ears. Thus, these action-word-object triads can either be consistent, namely that the same action accompanies successive presentations of a word-object association (i.e., the rabbit is always hopping), or inconsistent in that different actions accompany each presentation of a word-object association (i.e., the rabbit is sometimes hopping and sometimes wiggling his ears).

Indeed, consistency in the input to children appears to have a strong influence on early word learning. Consistent mappings of words and objects are obviously a prerequisite for word learning. This is best observed in cross-situational learning paradigms where learning is based on 12- to 14-month-olds tracking the frequency with which distinct words are presented with distinct objects across successive presentations (e.g., Smith & Yu, 2008). Moreover, studies have shown that consistency in referent location (Benitez & Smith, 2012, Samuelson et al., 2011), and consistent spatial, temporal and linguistic contexts (Roy et al., 2015) support word learning in early childhood. Also, studies on repetitive story telling find that 3-year-olds learned novel words better when being read the same book (containing the novel words) relative to being read different books with the same novel words (Horst et al., 2011, Williams & Horst, 2014). Finally, demonstrating the benefit of syntactic consistency, Schwab and Lew-Williams (Schwab & Lew-Williams, 2016) show that consistent sentence structure similarly helps 2-year-olds learn new words. Taken together, consistency plays a crucial role in the acquisition of early word-object mappings, impacting the strength with which such mappings are both formed and later accessed.

However, if more than two aspects in this context co-occur, the intended referents of the label may be ambiguous: for instance, if a particular action demonstra-

tion consistently accompanies presentations of a word-object association, the word could either refer to the action or to the object due to their similarly high rates of co-occurrence. If, however, the actions vary while the word and object are consistently presented together, the variability in the action presentation might help in disambiguating the object as the intended referent of the word.

Variability has typically been shown to play a role in structuring abstract knowledge, for example in categorization tasks. For example, variability in visual appearance across trials facilitated learning of perceptual categorization in infants as young as 6 to 7 months (Quinn & Bhatt, 2010). Also, 10-month-olds learned to discriminate between a typical and an atypical member of a category, only when the objects in that category varied strongly in their perceptual appearance, but not when they were perceptually similar (Mather & Plunkett, 2011). Further, hearing the same words produced by different speakers helped 14-month-olds to discriminate between the previously heard word and a similar sounding word, suggesting that variability in the phonetic detail of speakers may help a child to grasp subtle differences between those words (Rost & McMurray, 2009, 2010).

Twomey, Ranson, and Horst (Twomey et al., 2014) suggest, however, that there are limits to the effects of variability on learning: although 30-month-olds learned labels for categories when the objects varied in color (but not when they were in identical colors), children did not learn category labels when objects varied in shape *and* color. Thus, too much variability disrupted children's learning of category labels. In contrast, Junge, Cutler, and Hagoort (Junge, Cutler, & Hagoort, 2012) presented 9-month-olds with either constant picture-word pairings (i.e., always the same picture of a cat when hearing the label *cat*), or varying pictures of the same object (i.e., different cats when hearing *cat*). Both groups showed learning, indicating that neither consistency nor variability influenced their word recognition. Furthermore, Twomey, Ma, and Westermann (Twomey et al., 2017) tested 2-year-olds' word learning abilities while the background colors of the screen either varied

in color or stayed consistently white. They found that children only showed target recognition in the variable condition which led them to assume that decontextualization helped the children to form strong word representations. Nonetheless, they also suggest that increased variability might disrupt successful learning and that this might relate to the learning environment: reduced variability might help in rich learning environments and increased background variability may boost learning in simpler learning environments (see also Radesky & Christakis, 2016, Twomey et al., 2014). Thus, the effect of variability may vary across contexts, and guide attention allocation in different ways, thereby influencing learning behavior (see also Bahrick et al., 2002).

Importantly, as the literature reviewed above suggests, both consistency and variability in the input are required for learning because they seem to serve different functions as the literature reviewed above suggests: Consistency can help to form strong representations and to deepen your knowledge. In contrast, variability can help to categorise these representations and to broaden your knowledge. Accordingly, both aspects in combination can lead to consolidated and diverse knowledge of the world (Goldenberg & Sandhofer, 2013, Vlach & Sandhofer, 2011).

In the present study, we ask whether young children and adults better learn the labels of novel objects when particular actions *consistently* co-occur with the novel word-object pairings relative to when they are presented with *inconsistent* actions, i.e., when distinct word-object mappings are accompanied by varying action demonstrations.

We tested three groups of children, and adults as a control group. We chose 18-month-olds and 30-month-olds to capture children on either side of the vocabulary spurt. For the 18-month-olds, we expected that they would learn words for objects, but that their learning would be influenced by the consistency of actions. For the 30-month-olds, being past the vocabulary spurt, we expected that the consistency of actions would not influence their word learning, and that they would thus learn

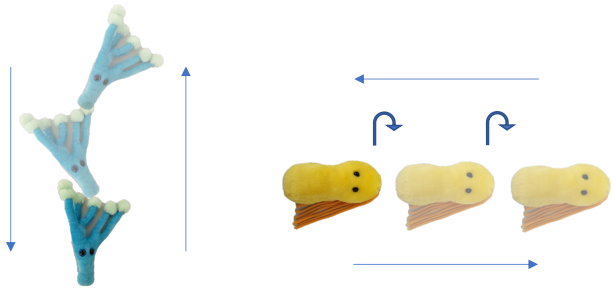
in both conditions. 3- to 4-year-olds were added on later following the results of the 30-month-olds, and we expected them to learn in both conditions as they are experienced word learners and might not be as easily distracted by actions. Therefore, this age group might be most similar to the adults whom we expected to learn in both conditions.

Methods

Participants. Fifty-four German monolingual 18-month-olds (range = 17;23 - 22;55 months; mean = 18;07, 25 girls) and sixty-five 30-month-olds (range = 27;09 - 32;12 months; mean = 29;72, 30 girls) and fifty-nine 3- to 4-year-olds (range = 38;89 - 47;9 months; mean = 43,44, 27 girls) participated in the experiment. In addition, 60 adults were tested as a control group (range = 19 - 50 years; mean = 26;32, 38 women). Twenty-seven additional participants were tested but excluded from the analysis because of unwillingness to participate (2), technical failure (2), calibration issues (4), bilingualism (2), preterm birth (2), impaired hearing (2), familiarity with the presented novel object (2), or insufficient data (11, see Preprocessing for details). Children were recruited from the group's database and participation was rewarded with a book. In order to control for children's language abilities, caregivers of 18- and 30-month-olds completed the German adaption of the FRAKIS (Fragebogen zur frühkindlichen Sprachentwicklung; Szagun et al., 2009). To test the language abilities of 3- to 4-year-olds we used the SETK3-5 (Sprachentwicklungstest für drei- bis fünfjährige Kinder; Grimm, Aktaş, & Frevert, 2010). Adults were mostly students of the University and were rewarded with either 0.5 course credit points or 4 €. Ethics approval was granted by the University of Göttingen (Project 123).

Stimuli. We selected two pseudo-words in keeping with German phonotactic constraints (Tanu and Löki), two arbitrary actions (explained in further detail below), and two novel objects (a yellow and a blue soft toy from `https://www.giantmicrobes.com/us/`, see Fig. 10). The auditory stimuli were recorded by

Figure 10. Blue and yellow toys were used as novel objects. As novel actions, an upward movement with leaning to the sides, and a sideways movement with tilting backwards and forwards were used.



a female German native speaker in infant-directed speech. The labels were embedded in a carrier phrase in both the training phase (e.g., "Schau mal, ein Tanu!", "Look, a Tanu!") and test phase (e.g., "Wo ist denn das Tanu?", "Where is the Tanu?").

The video stimuli consisted of training and test videos. In the training videos, participants saw a hand (with the arm of the agent being visible) moving the objects according to the two selected actions starting from the middle of the screen. We recorded separate videos of both actions being performed on both objects, such that across participants, videos of each distinct action being performed on each object were presented the same number of times. The first action involved a hand moving the object upwards while moving the object side-to-side (from left to right) with each increment in height. It started in the lower middle of the screen, went up and back down again. The second action involved a hand moving the object sideways, while tilting back and forth, moving first to one side of the screen, then back to the other, and ending up in the middle of the screen again. Both actions filled the whole 7 s of the trial presentation. The auditory stimuli, i.e., the labels for the objects embedded in carrier phrases, were presented at the same time to ensure temporal

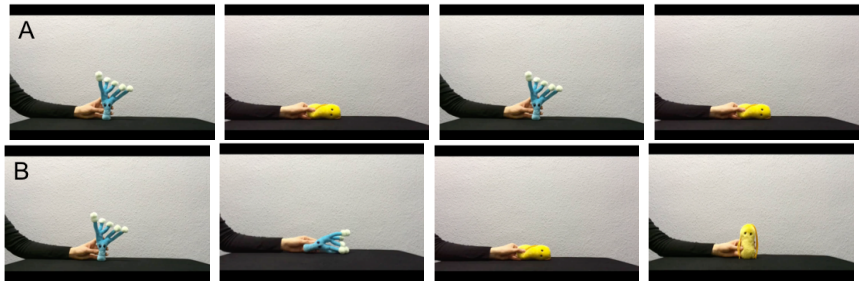
synchronicity of actions and language (Matatyaho-Bullaro et al., 2014). Each video in the training phase was 7 s long and 720 × 420 pixels in size.

Across the test trials, infants saw the two objects (still images) side-by-side on screen for 5 seconds and were led to fixate the target using carrier phrases (e.g., "Where is the Tanu?") such that the target label began 2.5 s into the video. Individual images of the two objects in the test phase each 640 × 480 pixels, and areas of interest for the analysis were defined accordingly.

Procedure. Participants sat in a dimly lit and quiet experimental room at a distance of 65 cm from a TV screen (92 × 50 cm). Children sat either in a car seat or on the parent's lap. A remote eye tracker (Tobii X 120), set on a platform underneath the TV screen, was used to record gaze data at 60 Hz. The software E-Prime was used to present the stimuli. Auditory stimuli were presented via two loudspeakers situated above the television screen. Two video cameras centered above the screen served to monitor the participant during the experiment. Calibration was conducted in Tobii Studio using a 5-point calibration procedure, in which a red point appeared in all four corners and the center of the screen. The experiment began following successful calibration. Each trial started with a Teletubby serving as a fixation cross in the middle of the screen against a black background, followed by the stimulus presentation. The videos in the training phase were presented in the center of the screen. Pictures in the test phase appeared next to each other in the center of the left and right half of the screen.

Experimental design. Each participant was presented with a yoked training and a test phase in a preferential looking paradigm (Hirsh-Pasek & Golinkoff, 1996). The training phase consisted of 4 trials, with each trial presenting the participant with the individual objects in motion accompanied by the label for this object (see Fig. 11). Participants were presented with 2 trials for each word-object association, and the pairing of objects with labels was counterbalanced across participants. Thus, *Tanu* was the blue object for half of the participants and the

Figure 11. Example of the training phase. In Panel A, each object is associated with one label and one action (Consistent condition). In Panel B, each object is associated with one label and both actions (Inconsistent condition). Each video in the training phase lasted 7 s, and an attention getter was presented before each trial. All videos are available on OSF under `osf.io/tndj7`.



yellow object for the other half. Participants were allocated randomly to one of two conditions (consistent or inconsistent) in a between-participants design. In the *Consistent* condition, they heard the labels for each individual object while seeing the same action being performed on this object (e.g., the blue object was always presented as moving up, counterbalanced across participants) across both trials. A second group of participants was allocated to the *Inconsistent* condition where they saw the same objects and heard the same labels. However, here, participants saw both actions being performed on each object across trials (e.g., the blue object was presented as moving up in one trial and moving to the side in a second trial).

The test phase consisted of 8 trials. Each trial presented participants with both images side-by-side on-screen as they heard the label for one of these images exactly half way through the trial, i.e., 2.5 seconds into the trial, dividing it into a pre- and post-naming phase (see Fig. 12). Gaze points to both images in the pre-naming phase provide an estimate of participants' baseline preference for either of the objects, while gaze points in the post-naming phase indicate participants' response to the relation between the heard label and the images on-screen (Hirsh-Pasek & Golinkoff, 1996).

Overall, the whole procedure lasted four to five minutes on average. Counterbalancing of novel words, actions, and objects resulted in 6 training lists. Additionally, 4 test lists were created (2 words $\times$ 2 objects, i.e., each object was paired with each

Figure 12. Test phase. Both objects are on-screen and the target is labelled at 2.5 s. The test phase consisted of eight trials with four trials per label. An attention getter was presented before each trial. All videos are available on OSF under `osf.io/tndj7`.

label across lists). Presentation order of all trials within blocks was fully randomized.

Preprocessing. The eye-tracker provides an estimate of where participants were looking at in each time stamp during the trial, with one data point approximately every 16 ms. All data (gaze data and trial information) were exported from E-Prime and then further processed in R (R version 3.2.4 (2016-03-10), R Core Team, 2016). For each time stamp, data were only included when one or both eyes of the participant were tracked reliably (validity less than 2 on E-Prime scale). When both eyes were tracked, the mean gaze point for both eyes was computed for further analysis. Gaze data were then aggregated into 40 ms bins. Areas of interest were defined according to the size of the displayed objects and their location on screen.

For the test phase, we coded whether the participant looked at the correct object (i.e., the target), the distractor or at neither of these images on-screen. We analyzed data that began 240 ms after the onset of the label presentation to ensure that we only analyzed eye movements that could be reliably interpreted as a response (Swingley et al., 1999, Von Holzen & Mani, 2012). Further, we subtracted the mean proportion of target looks of the baseline phase (PTL: total looking time at target/(total looking time at target and distractor)) at each 40 ms time point in the post-naming phase on that particular trial.

Single test trials were excluded if a participant looked at the stimuli more or less than M ± 2SD of the trial duration. Thus, a trial was excluded if an 18-month-old child looked 24.1% or less of the time during the trial to one of the two pictures. This led to an exclusion of 48 trials (10.5%). For 30 months, this criterion was at ≤ 25.7%, 62 trials (11.4%); for 3- to 4-year-olds it was at ≤ 40.6%, 40 trials (9%); and for adults at ≤ 37.6%, 28 trials (5.9%). This led to exclusions of some participants (included above) who did not look enough during any of their trials (18: $n = 3$; 30: $n = 3$; 36 - 48: $n = 1$; Adults: $n = 1$, see for example Andreu, Sanz-Torrent, & Trueswell, 2013 for similar values). Furthermore, participants were excluded from the analyses if they provided data from only one trial in the test phase (18: N = 1; 30: N = 1; 36 - 48: N = 1). This left us with 238 participants (n Consistent/n Inconsistent: 18: 28/26; 30: 33/32; 36 - 48: 30/29; Adults: 30/30).

Results

Here, we present two different analyses for each age group. In the first analysis, we report traditional ANOVAs and t-tests which evaluate our dependent measure, the proportion of fixations to the target collapsed across the post-naming window corrected for fixations to the target in the pre-naming window. In the second analysis, we report generalized linear mixed models. This allows us to include *Time* within the trial as an additional factor and examine how children's response develops across the course of the trial, since fixations change rapidly across the length of the time window tested in the current study. Using both of these approaches helps us understand the data from a more differentiated and situated perspective.

ANOVAs and t-tests. For a comparison of conditions across age groups, we ran a 2 (condition) × 4 (age) between-participants ANOVA. For each age group, we ran three t-tests: first a two samples t-test comparing the proportion of target looking in the Consistent condition and the Inconsistent condition. Then we ran separate one sample t-tests comparing the baseline-corrected proportion of fix-

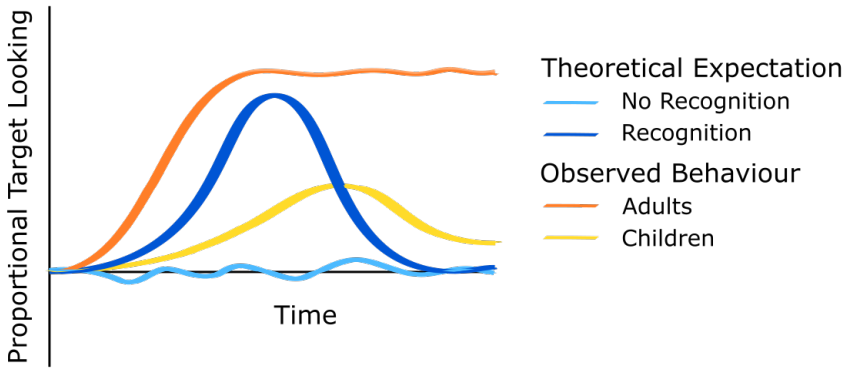
ations to the target in the post-naming window against chance (chance level = 0). Descriptives for each condition per age group can be seen in Tab. 6.

Table 6
Descriptives of baseline-corrected proportional target looking in the test phase for the Consistent and Inconsistent condition per age group (18 months, 30 months, 3 - 4 years, and adults). Scores of 0 reflect that averaged target looking is at chance level, i.e., no change from baseline), any values above 0 reflect target looking, and values below 0 reflect distractor looking.

Age group	Condition			
	Consistent		Inconsistent	
	Mean	SD	Mean	SD
18	-0.03	0.12	0.00	0.19
30	0.07	0.16	0.02	0.16
3 - 4	0.07	0.20	0.03	0.17
Adults	0.20	0.28	0.23	0.29

Description of the growth curve model. We also used Generalised Linear Mixed Models (GLMM) with *Time* as a factor to model children's looking behaviour during the test phase of the experiment (also coined Growth Curve Analysis (GCA), for a more detailed description and instruction see Mirman, 2014, Mirman et al., 2008). Including time, and its linear, quadratic and cubic polynomial allows us to capture any non-linear change in looking behavior across the whole trial duration (Mirman, 2014, Mirman et al., 2008). In Fig. 13, theoretical expectations and typical looking behaviour are plotted. If the participants do not recognise the target upon hearing the word (i.e., did not associate the word with the target), we expect target looking not to deviate from chance level across the whole time window, which would suggest that participants look more or less randomly at the two objects (light blue). If the participants do recognise the target upon hearing the word, we would expect a quadratic-shaped curve (or quartic if small tails are considered), reflecting how the participant is first at chance level, then looking more to the target upon hearing its label, and then going back to chance level (dark blue). In practice, adults usually stare at the correct object until the end of the trial (orange) (orange; see

Figure 13. Streamer plot of theoretical expectations and typical behavioural observations of looking behaviour across time, reprinted from Chapter 1.



also Mirman et al., 2008), which is often better fitted with a linear (steep rise) or cubic function (steep rise and high plateau). Children often show weaker effects in word learning studies with slower rises in the beginning and decreases towards chance level at the end of the trial (yellow), which can often be well-fitted with a cubic polynomial. However, children's target looking appears to be more variable in most cases, so flexibility in fitting the curve as well as in interpreting its course over time are necessary (e.g., Eiteljoerge et al., 2019b, Von Holzen, Fennell, & Mani, 2019). Based on these assumptions, we hypothesized that learning in any of the age groups would be reflected on the quadratic or cubic polynomial, and potentially on the linear term for the adults. If then, participants learn in one condition but not the other, we expect the factor Condition to interact with the respective polynomials which would reflect different curves for the two conditions.

We fitted a generalized mixed model using lme4's lmer function in R (Bates et al., 2015) with Gaussian error structure and identity link function. Condition and Time, and their interaction were included as fixed effects of interest. Further, Label (Tanu, Löki), Object color (blue object, yellow object), and Age at test were included as control factors. Age at test was included in addition to the age group, as the oldest group of children had a larger range in age from three to four years. We also included Participant id as a random factor to allow for random slopes across participants (Mirman, 2014). Age at test was z-transformed (to a mean of zero and

a standard deviation of one) to simplify later interpretation. A reduced model was fit that did not include Condition. A comparison between the reduced model and the full model including the factor Condition then allows to evaluate the influence of the factor Condition.

As noted above, we also included *Time* and its second and third polynomial in the model. This allowed us to model the data as a linear, quadratic, and cubic function of time, and thus, to capture according curvatures in the data (the number of polynomials has been restricted to three based on theoretical considerations regarding looking behavior). An overview of the results can be seen in Tab. 12 and 13. We used the function drop1 to evaluate the influence of each factor in the model. This function compares the model including one factor with a model without this factor, and thereby evaluates its contribution to the model.

Visual inspection of a qq-plot and a histogram of the residuals showed a normal distribution, but homogeneity appeared to be violated based on a plot of residuals against fitted values. Log-transforming the response did not contribute to an improvement of the model. We, therefore, chose the first model, but results need to be reviewed with care.

All age groups.

ANOVA. For a comparison of conditions across age groups, we ran a 2 (condition) $\times$ 4 (age) ANOVA. We found a significant effect of Age ($F(3, 230) = 13.65$, $p < .001$, $\eta_p^2 = .15$), but no main effect of Condition or an interaction of the two factors (ps $> .7$). These results suggest that there were differences among age groups but not across conditions or between conditions within age groups.

GLMM. Comparing a model including Condition with the reduced model without Age or Condition revealed that Condition did not significantly improve the model ($\chi^2 = 3.49$, df $= 4$, $p = .48$). However, including Condition and Age in the interaction did improve the model significantly compared to a Condition-only model ($\chi^2 = 64.06$, df $= 24$, $p < .001$) or an Age-only model ($\chi^2 = 27.6$, df $= 16$, $p = .035$).

Using drop1, the model revealed a significant interaction of Condition*Age*poly1 (χ^2 = 8.68, df = 3, p = .034) and Condition*Age*poly2 (χ^2 = 9.97, df = 3, p = .019), suggesting differences between age groups and conditions on the linear and the quadratic term. Thus, these results reflect differences in the time course of target recognition between the two conditions across age groups.

To analyze these effects further, we split the data according to the different age groups. To ensure that the participants truly learned the words in both conditions in the age groups, we further split the data according to condition. The results can be seen in Tab. 12 and 13[2].

18-month-olds.

T-tests. For the 18-month-olds, a Welch two sample t-test comparing Consistent and Inconsistent conditions found no significant difference between conditions, p = .57. Separate one sample t-tests comparing baseline-corrected target fixations in each of the conditions to chance level (chance = 0) were not significant (all ps > .1).

GLMM. The model comparison between the full and the reduced model was not significant (χ^2 = 1.16, df = 4, p = .89). When using drop1, no interactions between the conditions and the time terms could be observed. This suggests that Condition did not interact with 18-month-olds' target recognition.

For neither the model examining the data from the *Consistent* group alone (*Consistent* split model), nor the model examining the data from the *Inconsistent* group alone (*Inconsistent* split model), did we find significant interactions with the polynomials. Thus, there was no evidence that 18-month-olds had learned and later recognized the word-object mappings in either condition.

[2]Running the same models with a subset of the participants who provided vocabulary information showed that vocabulary knowledge did not influence the results significantly. For comparability across age groups and to preserve a bigger sample size, we focus here on the models without vocabulary information.

30-month-olds.

T-tests. For the 30-month-olds, a Welch two sample t-test comparing Consistent and Inconsistent conditions found no significant difference between conditions, $p = .22$. Separate one sample t-tests comparing baseline-corrected target fixations in each of the conditions to chance level (chance = 0) showed a significant effect in the Consistent condition (t(32) = 2.69, $p = .011$, d = 0.66), but not in the Inconsistent condition ($p = .4$).

GLMM. The model comparison between the full and the reduced model was above the threshold of significance we adopted in the current study ($\chi^2 = 9.26$, df = 4, $p = .055$). When using drop1, the model revealed a significant interaction of Condition*poly2 ($\chi^2 = 4.99$, df = 1, $p = .025$). This result suggests that there are differences in 30-month-olds' target recognition across conditions on the quadratic term.

For the *Consistent* split model, the model revealed a significant effect on poly2 ($\chi^2 = 6.8$, df = 1, $p = .009$). For the *Inconsistent* split model, we did not find any significant influences of the time terms. Thus, 30-month-olds had learned and later recognized the word-object mappings in the consistent, but not in the inconsistent condition.

3- to 4-year-olds.

T-tests. For the 3- to 4-year-olds, a Welch two sample t-test comparing Consistent and Inconsistent conditions found no significant difference between conditions, $p = .69$. Separate one sample t-tests comparing baseline-corrected target fixations in each of the conditions to chance level (chance = 0) showed no significant effects in either condition (all ps $> .1$).

GLMM. The model comparison between the full and the reduced model was significant ($\chi^2 = 13.26$, df = 4, $p = .01$). Using drop1, there was an interaction between Condition*poly1 ($\chi^2 = 5.36$, df = 1, $p = .021$) and the interaction of Condition*poly3 was above the threshold of significance adopted in the current

study ($\chi^2 = 3.77$, df = 1, $p = .052$). These results suggest there was a difference in the pattern of 3- to 4-year-olds' target fixations across the two conditions. Visual inspection revealed that in the *Consistent* condition, children showed a steep rise in fixating the target shortly after the onset of the target label (from around 500 ms) with a peak around 4250 ms. Contrary, looks to the target in the *Inconsistent* condition hovered at (or just above) chance throughout the trial (see Fig. 14).

For the *Consistent* split model, the linear time term was significant ($\chi^2 = 5.92$, df = 1, $p = .015$) and the cubic term was above the threshold of significance adopted in the current study ($\chi^2 = 3.48$, df = 1, $p = .062$), suggesting that 3- to 4-year-olds had learned the word-object mappings presented to them in the training phase in the *Consistent* condition, reflected in the sharp increase in their target looking.

For the *Inconsistent* split model, we did not find any effect on the polynomials. Thus, there was no evidence that 3- to 4-year-olds learned the word-object mapping in the *Inconsistent* condition.

These results highlight the differences in word learning success between the *Consistent* and *Inconsistent* condition at 3- to 4-years of age: target recognition followed different trajectories in the two conditions which is reflected in the linear time term. The split models reveal that only in the *Consistent* condition, did children learn the word-object associations. This effect was not evident in the t-tests, probably due to the late increase in target looking in the consistent condition which resulted in a lower overall mean.

Adults.

T-tests. For the adults, a Welch two sample t-test comparing Consistent and Inconsistent conditions found no significant difference between conditions, $p = .81$. Separate one sample t-tests comparing baseline-corrected target fixations in each of the conditions to chance level (chance = 0) showed a significant effect in the Consistent condition (t(29) = 4.02, $p < .001$, d = 1.04), and in the Inconsistent condition (t(29) = 4.16, $p < .001$, d = 1.07).

GLMM. The model comparison between the full and the reduced model was not significant ($\chi^2 = 4.44$, df = 4, $p = .35$). Using drop1, the model revealed no significant interactions of condition with the time terms. Thus, these results suggest that target looking across time did not differ between conditions for the adults.

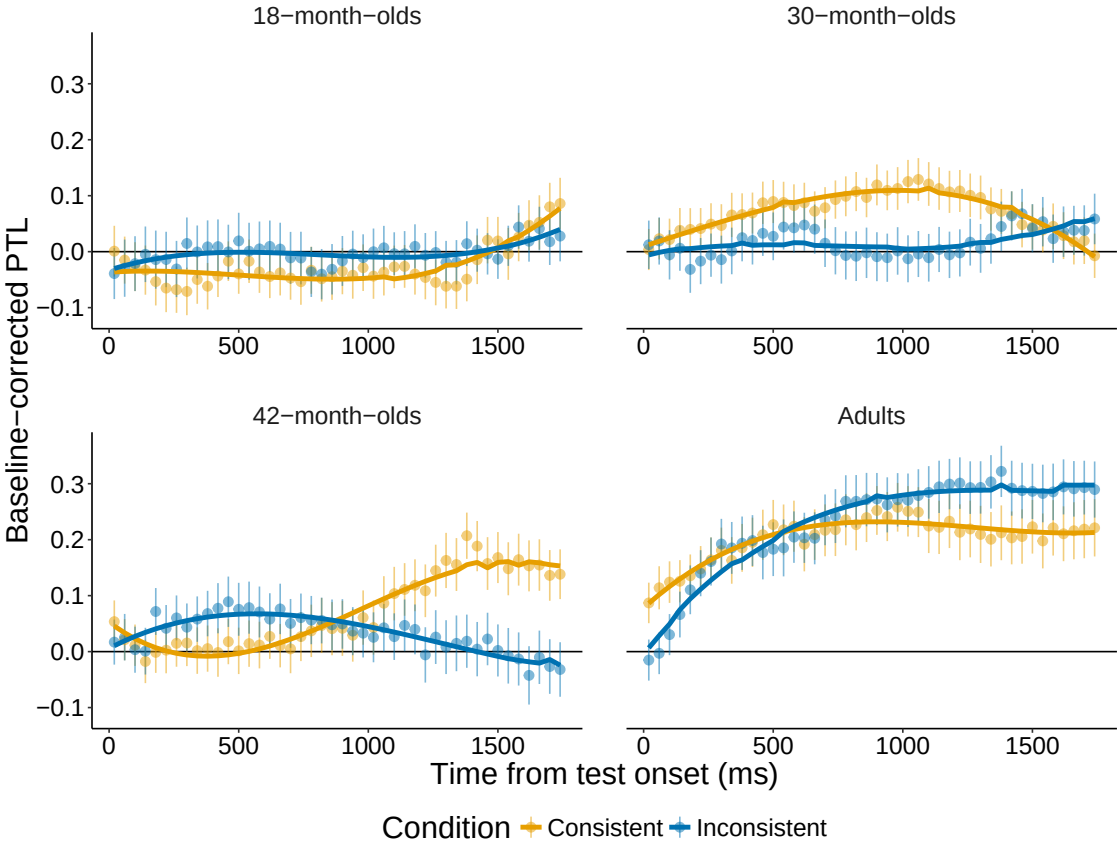
For the *Consistent* split model, the model revealed a significant effect on poly2 ($\chi^2 = 5.59$, df = 1, $p = .018$). For the *Inconsistent* split model, the model revealed a significant effect on poly1 ($\chi^2 = 13.8$, df = 1, $p < .001$), and poly2 ($\chi^2 = 9.79$, df = 1, $p = .002$).

These results suggest that adults recognized the target successfully in both the consistent and the inconsistent condition. Even for the adults, target recognition followed slightly different trajectories in the two conditions, as can also be seen in Fig. 14. In the *Consistent* condition, fixations to the target show a steeper rise and a lower and earlier plateau compared to the *Inconsistent* condition. However, this difference was not significant in both the t-test and the model comparison.

Discussion

In the current study, we investigated whether the consistency with which particular actions co-occur with particular word-object pairings influences children's word learning. Participants were presented with novel labels for novel objects while these objects either moved in a *consistent* (i.e., always the same action across different trials presenting the word-object mapping) or in an *inconsistent* (i.e., both actions performed on each object across the different trials presenting the word-object mapping) manner. We did not find any evidence that 18-month-olds learned the word-object mappings in either condition. In contrast, 30-month-olds and 3- to 4-year-olds learned the word-object mappings only in the *Consistent* condition. This was reflected in the linear, quadratic, and cubic time terms of the GLMM, representing the rise and fall of target looking over time in this condition. Only adults learned words for objects in both conditions.

Figure 14. Time course graphs for each age group of participants' baseline-corrected proportional target looking (PTL) during the test phase after label onset at 2500 ms and 240 ms to initiate a fixation. The *Consistent* condition is represented in yellow, the *Inconsistent* condition is in blue. The line at 0 represents chance level, increases reflect proportionally more target looking whereas decreases reflect distractor looking. The yellow and blue lines reflect the fitted GLMM including Time up to the cubic term for each condition.



These results suggest that the consistency of co-occurring actions influences 30-month-olds' and 3- to 4-year-old children's word learning. As Gogate et al. (Gogate et al., 2000) have shown, actions and words often co-occur in the child's multimodal environment. These actions have also been shown to support word learning (Werker et al., 1998), especially when provided with temporal synchronicity (Matatyaho-Bullaro et al., 2014). However, if these actions appear to be referential, 15-month-olds' word learning is exacerbated (Puccini & Liszkowski, 2012). Our results extend these findings in a critical way, showing that it is not just the temporal consistency with which actions accompany word-object associations that impacts word learning success. Between 30 months and 3- to 4-years of age, children

learned the word-object mappings only when each word-object mapping had previously been presented with the same action being performed on the object across separate presentations, i.e., in the consistent condition. This benefit of consistency was not observable at younger ages or with adults. We interpret these findings to suggest that redundant information from different domains supports the forming of rich lexical representations, but only if this information highlights the word-object association and does not distract from it, and only towards the third year of life.

This is in line with the literature suggesting a beneficial effect of consistency on word learning (Roy et al., 2015), in contrast to a beneficial effect of variability on generalization over different members of a category (Ankowski, Vlach, & Sandhofer, 2012, Perry et al., 2010, Twomey et al., 2014). Nevertheless, some recent work suggests an impact of lower-level variability (variability in the colour of the background on which objects were presented) on learning of word-object associations (Twomey et al., 2017). In contrast to the beneficial effect of variability in the study by Twomey et al., no such effect was found in our study. We suggest that the function and salience of the additionally varying information might play a crucial role in whether it boosts or detracts attention to word learning. In Twomey et al.'s study, the background colors were a subtle manipulation of the variability in the stimuli. Actions performed on objects, in contrast, may be more salient and compete for attention with the word-object mappings leading to children ignoring the word-object mapping in favor of increased attention to the actions presented. Increased and salient variability can, therefore, disrupt successful learning in a complex learning environment (see Radesky & Christakis, 2016, Twomey et al., 2014). The results of the current study, thus, highlight the importance of a developmental perspective on the influence of cross-domain information on processing. Taken together with the literature on intersensory redundancy, we trace here a developmental pattern in the influence of actions on word learning. Early in life, with actions that do not detract attention from the linguistic input but rather highlight the association between the

linguistic input and objects in the world, co-occurring actions bolster word-object association learning. With increasing age, potentially increasing salience of actions (18 months), and variability in the actions presented (30 months, 3- to 4-years), actions may indeed deter from word learning and lead to children failing to learn the intended word-object mapping. Ongoing work (described in detail below) helps clarify this picture with regard to the salience of actions and words across development.

Dynamic systems and emergentist approaches to development assume that language acquisition and action processing take place in a complex environment where no single aspect of the environment is *a priori* responsible for development in any domain (Hirsh-Pasek et al., 2000, Smith & Thelen, 1993, see Smith & Thelen, 2003, for an overview). Within this approach, development is the product of active interaction between the child and her immediate environment. As the child develops, she learns to integrate increasing amounts of information while the direct environment adapts to the child's needs and provides the opportunity to learn in a rich and multimodal world.

In their adaptation of such accounts focusing on early word comprehension, Gogate, Walker-Andrews, and Bahrick (Gogate, Walker-Andrews, & Bahrick, 2001) suggest that word learning results from the interaction of several processes, including selective attention and intersensory perception, as well as the multimodal input provided to the child (based on J. W. Sullivan & Horowitz, 1983, and their work on early intermodal perception). Information from different sources and senses regarding the same object might seem unrelated at first, but, when these information sources co-occur, they help to form a full-blown representation. For example, a word and an object might seem arbitrary when presented independently. But presented in temporal synchrony (Gogate et al., 2000) and potentially even with movement of the object (Gogate & Bahrick, 1998, Werker et al., 1998), the relationship between the word and the object is highlighted for the infant, and therefore supports learning.

Importantly, perceiving and encoding the relevant aspects of the input requires that the child attends to this input (and indeed, to the relevant aspects of this input). However, attention allocation might be guided by the child's momentary interest in different features of the input. Thus, focusing on a certain type of information can lead to impoverished encoding of other information, which in turn, will shape the spiral staircase of learning in a particular way. Therefore, information from different domains can and does impact processing and learning in the different domains, and can either help or hinder learning across development: Only when the child is able to encode all relevant aspects of the complex multimodal input provided to her, will she be able to avoid such momentary pitfalls of attention allocation. Otherwise, learning can be impeded by the complexity of the input that is beyond the child's resources of attention and perception.

With regard to such dynamic accounts of learning, our results suggest that the multimodal complexity of the input interacts with the child's abilities to learn. Multimodal input appears to be helpful in certain learning contexts (e.g., Matatyaho & Gogate, 2008), but a child will only learn what is perceptually salient and relevant in his or her contextual and developmental situation at that specific moment (e.g., Gogate et al., 2001). Importantly, we find that the presentation of multimodal input has an impact on word learning, even though this influence is detrimental to word learning at younger ages, especially when co-occurring with salient information from the non-linguistic domain. Our results highlight how children's perception, depending on the age and the developmental context of the child, can be challenged by presenting salient variability in a multimodal context (Twomey et al., 2017), and how this can directly influence learning: words might not be learned as easily if the child is provided with a learning environment that is highly complex, or if that learning environment provides other information that might attract the child's attention. Until age 4, action consistency seems to have a positive cross-domain influence on word learning. In contrast, adults seem robust against the cross-modal

influences between word and action learning. Thus, by adulthood we find limited influence of other domain information on word learning.

We note, however, that we cannot conclude from the current study that children devoted more attention to the actions relative to the word-object mappings, since children were presented with words and actions simultaneously (Matatyaho & Gogate, 2008). A related study in our lab finds, in keeping with these findings, that 12-month-olds do not learn word-object mappings when these are accompanied by actions (see also Puccini & Liszkowski, 2012), while learning to associate actions with objects. Older children (24 and 36 months) do learn word-object mappings but action-object mappings are less successfully learned, suggesting that, at the age where we find an effect of action consistency on word learning, children are able to successfully learn words for objects (Eiteljoerge et al., 2019b). Thus, from a certain age, auditorily presented words may be more salient compared to visually presented actions. And, while being influenced by the consistency of visually presented actions, children are not distracted by these to the extent that they fail to learn words for objects. However, for the 18-month-olds, we did not find learning in either condition. In accordance with the other study in our lab, this could suggest that the visually presented actions here attracted the children's attention to an extent that hindered word learning in the present study. These findings are in contrast to the literature on auditory overshadowing: In these studies, young children usually show an auditory preference, while older children vary in this preference and adults often show a visual preference (see Barnhart et al., 2018, Robinson & Sloutsky, 2004, for shifts in modality dominance). This developmental trajectory would suggest, in contrast to our suggestions above, that young children focus on auditory words before focussing on visual actions. Thus, our results are difficult to reconcile with those findings and would suggest that auditory dominance develops after the first year. However, the differences in the results could also be due to the types of stimuli presented: object-manipulating actions like the ones used in the present study

might appear very salient in contrast to the still images often used in the modality dominance studies (Barnhart et al., 2018).

We note also that our findings with 18-month-olds stand in contrast to previous work with younger infants (Werker et al., 1998) in which even 14-month-olds learned words when these words were presented with objects in motion. In this study, one object (i.e., a dog) was always presented moving to the front of the screen and back while the other (i.e., a toy truck) was presented moving from one side of the screen to the other. Here children only learned the word-object mappings when the objects were presented in motion but not when they were presented without motion,. Notably here, the 14-month-olds were presumably already familiar with images of the dog and the truck and only had to learn the mapping between the novel object and its label, and it is likely that this increased familiarity with the objects may have led to children mapping the words onto these objects with greater ease (Fennell & Waxman, 2010). The salience of the familiar objects may, therefore, have overridden the effect of salience of the actions on word learning. Furthermore, we note that in some cases, e.g., the study by Gogate and colleagues, infants were presented with 3D objects in a live interaction with their caregivers (Matatyaho & Gogate, 2008). This contrasts with our passive screen-based setting which might have made it more difficult to learn the words (see studies on the video deficit, e.g., Kuhl, 2007, Roseberry, Hirsh-Pasek, & Golinkoff, 2014).

Finally, with regard to the two different statistical approaches, we believe that combining both approaches helped us to quantify the effects we observed. ANOVAs and t-tests have been used more often than GLMMs in the word learning literature, and using this approach therefore allowed us to compare our results to previous results in the literature. At the same time, GLMMs which include *time* as a factor provide higher temporal resolution and allow us to include other factors which might be responsible for variance in the data. For example, the GLMMs showed a quartic curve for 18-month-olds' word-object association learning in the consistent

condition, which was absent when target looking was averaged across time. Furthermore, the quartic curve was probably due to a subtle increase towards the end of the trial, and is therefore difficult to reconcile with the theoretical assumptions outlined above. Furthermore, the GLMMs showed a curve for 30-month-olds' word learning in the consistent condition, and again, this effect was absent when target looking was averaged across time. These differences in results of the 30-month-olds suggest that they did learn the word-object associations in this condition (since target looking followed a predicted curve), but their target looking was only observable when we considered time during the trial to allow for changes in the pattern of looking behaviour across the trial. Even for adults, although they showed learning in both conditions across statistical approaches, the pattern of target looking differed across time during the trial, which might suggest differences in processing. Thus, the combination of ANOVAs, t-tests, and GLMMs allowed us to present a more differentiated picture of the participants' behaviour. Nonetheless, more research is required with regard to the GLMM and the interpretation of different types of curves.

In conclusion, we find that children's word-object learning between 30 months and 4 years of age was influenced by the consistency with which particular actions co-occur with word-object pairings: in this age range, *consistent* word-action-object mappings supported word-object learning in contrast to *inconsistent* word-action-object mappings. Only adults learned words independent of the actions in the present setting. In terms of a dynamic systems account, this development reflects how the interaction between the learner and the complexity of the multimodal environment plays a role in shaping the learning experience: It shows how we learn to incorporate parts of a rich multimodal environment that learners of any age experience.

Discussion

Already in early childhood, children are confronted with a complex environment from which they need to extract information to learn about the world. For example, caregivers' play encompasses both words and actions for objects. But which associations do children learn in such a scenario? And which factors influence learning? The aim of the current thesis was to shed light on children's word and action learning, and the interactive nature between the developing child and the environment during learning. We investigated whether word-object and action-object learning develop in synchrony across early childhood, whether the child's own preferences contribute to her learning behaviour, and whether the consistency and variability in the co-occurrence of actions in the input influences the child's word-object learning across development.

In the first study, presented in Chapter 2, we examined to what extent young children and adults choose to associate words or actions with objects when presented with both simultaneously. All groups were presented with two objects which were associated with a distinct label and a distinct action. In the following test phase, both objects were presented on screen while participants' target looking was measured. In a prime phase, either the label (word-object trials) or the action (action-object trials) was presented, followed by the same display of the two objects. If participants looked more to the target than to the distractor object this was taken to be suggestive of successfully mapping either the word or the action with the object. 12-month-olds learned action-object associations, 24-month-learned neither word- nor action-object associations, 36-month-olds learned word-object associations, and adults learned to associate both words and actions with objects. Thus, word and action learning followed different trajectories in early childhood.

In the second study, presented in Chapter 3, we were interested in how preferences interact with children's learning of words and actions for objects. Therefore, we designed a gaze-contingent 2-choice paradigm in which participants could choose

either a word or an action for an object. In a first preference phase, participants were presented with familiar objects and according labels and actions. In the following exposure phase, we used the same paradigm, but this time, participants were presented with novel objects, labels, and actions. The test phase was the same as in Study 1. Children between two and three years learned words but not actions for objects, and preference influenced their time course of target looking in the word-object but not in the action-object condition. Furthermore, their preference and exposure choices were positively correlated, suggesting that their choices in both phases followed a similar strategy. Adults learned both words and actions for objects, and preferences modulated target looking in the action-object condition. Their choices in the preference and the exposure phase were not correlated. Thus, we replicated the results of Study 1, and found additionally, that preferences influenced processing of information.

In the third study, presented in Chapter 4, we investigated how the consistency in the co-occurrence of actions can influence young children's word learning. One group of participants always saw the same action for an object while it was labelled (consistent condition). Another group saw different actions for an object while it was labelled (inconsistent condition). 18-month-olds did not show learning in either condition at test, and 30-month-olds and 3- to 4-year-olds showed word learning in the consistent but not in the inconsistent condition. Adults learned words for objects equally well in both conditions. Thus, the structure of the multimodal information in the input influenced learning in toddlers but not adults.

Together, these studies show that children in the present studies developed from learning action-object to word-object to both associations, preferences influenced children's target looking in the word-object condition and adults' target looking in the action-object condition, and consistency in co-occurring actions helped children learn words for objects between 30 and 48 months, but was irrelevant for adults'

success in word learning. In the following, I first discuss these results in the light of the current literature, and then turn to their theoretical implications.

Parallels and differences in word and action learning

In the literature to-date, most research has shown that children from the age of 30 months show similarities across word and action learning, for example, in the extension of novel information to novel objects (Childers & Tomasello, 2003), referent selection through exclusivity (Riggs et al., 2015), and attention to novelty (Dysart et al., 2016). We found that children around 12 months learned actions for objects and, in Study 1 and Study 2, children between two and three years of age learned words for objects. Thus, while the current literature suggests that children show similarities in their word and action learning above the age of 30 months, our results show differences in their word and action learning below the age of 36 months, and further suggest different developmental trajectories across the two domains.

One crucial difference between the previous studies in the literature and the studies presented in this thesis is the setting in which children learn: in most previous studies, children learned either words or actions for objects, and therefore encountered the information in a unimodal context (e.g., Childers & Tomasello, 2002). In the present studies, we tried to mimic caregivers' multimodal interaction with children (Gogate et al., 2000), and therefore presented both words and actions simultaneously (albeit temporal synchrony in Study 1 and Study 2). Accordingly, in our bimodal setting, the two domains might have competed for the child's attention and processing abilities which resulted in children learning only in one domain. However, another study in our lab suggests that also in unimodal settings, we find different trajectories in early childhood. In this study, we presented children with the same words and actions for objects as in Study 1 and Study 2, but this time, we presented either words or actions for objects. We replicated the findings from

Study 1, showing that 12-month-olds learned actions for objects while 24-month-olds learned neither words nor actions for objects (Eiteljoerge, Adam, Elsner, & Mani, in prep.b). Thus, also in a unimodal setting, where competition of words and actions was avoided, children showed the same learning pattern as in our previous studies.

Another possible explanation for the differences between previous and our studies could be that children's word and action learning align with time when children become more flexible in processing and integrating new information. This assumption would predict that, in early years, the two domains follow different trajectories, whereas later in development when children possess more developed processing abilities, the two domains follow similar trajectories. This idea is supported by the study by Hahn and Gershkoff-Stowe (2010) with 4- to 5-year-olds and our results with adults in Study 1 and Study 2: participants successfully learn in both domains at later ages, and this suggests that the learning behaviour changes with age and the developing abilities of the child. Thus, children's differences in word and action learning in the present studies might be related to their developing abilities in early childhood, and similarities in learning between the two domains can be captured at later ages when children's processing abilities align with the task at hand.

The influence of learners' preferences

According to Montessori (1914) among others, the child as the learner seeks a learning context in which her own abilities and interests align with the demands and contents of the input. In Study 2, children showed individual preferences for words and actions. The correlation of their choices in preference and exposure phase might reflect how children's interest correlated with their information seeking behaviour when presented with new information. Whether this relation can also be observed causally, meaning that children's preferences guide their information seeking, cannot be inferred from these results. However, previous research suggests

that the relation between them might be bidirectional: curiosity seems to influence information seeking (Gottlieb et al., 2013), and gaps in our current knowledge seem to induce curiosity (e.g., Wade & Kidd, 2019). The absence of a correlation between preference and exposure in adults in Study 2 might suggest that adults change their choice behaviour given the type of information (i.e., familiar or novel), and thereby optimize their information sampling behaviour. However, since children and adults were not sensitive to the side violation in Study 2, we cannot be sure that participants associated one type of information with one side of the screen, and made their choices accordingly. Thus, from the current study, it is unclear whether their choices really indicated a preference for that domain. For children, the positive correlation of preference and exposure would also be compatible with alternative explanations, such as selective attention or individual side preferences. For adults, this correlation is difficult to interpret given its null result. To clarify this aspect, the relation between gaze-contingent designs, active choice, and preference requires more investigation in the future.

The results of Study 2 indicate that children's target looking in the word-object condition and adults' target looking in the action-object condition were influenced by individuals' preferences for the respective domains. In both groups, a higher preference for the domain revealed a pattern that was theoretically expected while a lower preference was related to similar, but delayed patterns of target looking. Our results also show that a preference for a domain is neither crucial nor sufficient for learning: On the one hand, both age groups learned in the domains where preferences influenced target looking showing that preferences modulated target looking but did not disrupt learning. On the other hand, a high action preference was not correlated with better action-object learning at two to three years of age suggesting that learning does not automatically follow individuals' preferences (see also Wade & Kidd, 2019). This indicates that additional factors, potentially the

form of the input and the child's processing abilities among others, contribute to learning in this situation.

The findings from Study 2 might help explain the absence of learning at 24 months in Study 1: it could be that children's preferences influence their information processing in this study, and this influence might be especially dominant around the age of two when processing abilities are still limited: In this case, their preferences could help to extract information from the input by reducing the focus of attention to subjectively relevant information. On the group level however, learning of such information might not be observable because children with different preferences focus on different aspects of the input, and thus, cancel each other out when we evaluate learning without taking preferences into account.

Also in Study 3, children's individual preferences might have influenced the results. If children showed an interest in words, the manipulation of co-occurring actions might be less dominant in their perception of the input, and thus, might only have a limited influence on children's word-object learning. In contrast, if children showed an interest in actions, this manipulation might be highly salient in their perception of the input, and thereby impede word-object learning accordingly. There is no overall preference for words or actions between the age of two and three years in Study 2, and therefore, we cannot conclude that children around 30 months in Study 3 had a certain preference on the group level that influenced their learning behaviour at the group level. Accordingly, it would be interesting to investigate interactions between the perceived salience of the input (potentially mediated by children's domain preferences) and the consistency and variability that is provided in the input. Such research could inform our understanding of how consistency and variability contribute to learning while taking the learner's perspective into account.

Consistency and variability

In addition to the child's own characteristics, the provided environment can influence children's learning behaviour. We found that children between 30 months and four years learned words for objects only when the word-object mapping co-occurred with the same action, and not when they co-occurred with varying actions across trials. Adults learned words for objects independent of the presented actions. These results are in accordance with the idea that consistency helps to deepen knowledge: in our context, the consistency in co-occurring actions helped learning a word for an object. This influence faded with age and increased processing abilities as indicated by the results of the adults.

As argued in the introduction, a possible alternative hypothesis was that consistency impedes learning in contrast to variability: The statistically high co-occurrence of information between words, actions, and objects in the consistent condition could have led participants to insecurities given the different possible mappings. In contrast, the lower statistical rate in co-occurrence in the inconsistent condition could have helped participants to extract the word-object mapping. Since this was not the case, it could be that children did not experience uncertainty in mappings, and potentially mapped both word and action with the object, or that certain cues guided them in their word-object learning. For example, the syntactic structure of the sentence might have sufficed for them to identify the target word as a noun and therefore predominantly map it onto the object (and not as a verb mapped to the action). The results of Study 1 suggest that while younger children might hold different hypotheses for potential mappings, this was not the case for older children. Accordingly, other cues such as the syntactic structure might have helped the children above the age of 30 months to learn the word-object mapping in the consistent condition in Study 3. Note however, that we did not test for their action-object or action-word association in this study, and therefore, cannot exclude the possibility

that they held all possible mappings as potential candidates, and showed learning of the word-object association at test given that this was the only one tested.

Consistency was also provided in Study 1 and Study 2 in form of the same word-action-object mappings across trials. Studies looking at cross-situational word learning, for example, use the consistency in mappings across encounters as the only cue to word-object associations, and suggest that children from the age of 12 months can use this high co-occurrence rate to determine the referent of a word (Smith & Yu, 2008). If we assumed that consistency in mappings is such a strong cue that it leads to learning, this would mean that children should have learned both words and actions for objects in Study 1 and Study 2. However, the more differentiated results in both studies suggest that consistency is one factor that contributes to learning, but the child's own developing abilities and preferences also contribute to the acquisition of knowledge.

Variability in the input complements consistency during learning because it can help to broaden knowledge (Goldenberg & Sandhofer, 2013, Vlach & Sandhofer, 2011). While it was not necessary to broaden knowledge in the current studies, the variability in Study 3 could have helped to learn words for objects due to the higher co-occurrence rate of word-object compared to action-object associations in the inconsistent condition. However, since children did not show learning in this condition, it might be that the variability in actions led the attention away from the word-object association, and thereby impeded learning. This contrasts with the findings by Twomey et al. (2017) which suggest that variability can support word learning. As discussed in Study 3, we suggest that the salience of such variability modulates learning: while the manipulation of objects might be high in salience, and therefore potentially distracting, more subtle input variability, such as the change in background colours, might help sustain attention, and therefore support learning (see also Radesky & Christakis, 2016, Twomey et al., 2014). To evaluate

this hypothesis, future studies need to investigate how different degrees of salience in the input interact with children's learning in consistent and variable contexts.

The differences in learning in the context of consistency and variability might also be influenced by the developmental trajectory of word and action learning, as found in Study 1 and Study 2: if young children's attention is drawn to actions, as the 12-month-olds' results in Study 1 suggest, learning words in a situation where the action of the object changes with each trial might be more difficult because the child might attend more to the changing actions, and therefore, focus less on the word-object mapping. This behaviour might change with development when children begin to focus on words rather than actions for objects. Following this train of thought, we could predict that children around three years are less influenced by co-occurring actions in the input when learning words because, according to Study 1, 3-year-olds focus on learning words and not actions for objects. However, even at three to four years, we find that children learn words only when provided with consistently co-occurring actions. Nonetheless, the results of the action violation phase in Study 2 show that children between two and three years of age detect a mismatch between actions, suggesting that they still process the actions on-display although they do not retain that information for longer, as the absence of target recognition in the test phase suggests. Thus, also for older children, varying actions might still pose high processing demands on the child, and therefore, impede word-object association learning.

Together, the three studies reflect how children and their abilities, and the environment and its characteristics interact in word and action learning during early childhood. Before I turn to the theoretical implications of the findings, I will shortly discuss the employed methods and statistics.

Methods and statistics

While Study 1 and Study 3 relied on the well-established intermodal preferential looking paradigm (Hirsh-Pasek & Golinkoff, 1996), we extended this paradigm to account for participants' preferences by using a gaze-contingent design in Study 2. This paradigm allowed us to incorporate an interactive part in which participants were actively engaged in the task within the limits of our setting. Thus, to a certain extent, such a paradigm can provide a bridge between purely observational and purely predefined and highly controlled study designs. However, although the paradigm is well-established in the literature on reading behaviour (Rayner, 2014), it is less well researched in younger populations, as is evident by the limited number of existing studies. Only recently, this paradigm has gained in popularity, and because of its recency, there are few established standards for younger age groups so far. One example is the length of the fixation that is required to initiate the gaze-contingent element. In our study, we chose 1000 ms because we wanted to allow for a decision making process. In contrast, other studies have used 400 ms (Vernetti et al., 2017), 150 ms (Deligianni et al., 2011), 100 ms (Vernetti et al., 2018), or the first fixation (Wang et al., 2012), and it is not clear whether and how these differences in fixation time can influence participants' behaviour in the task or which duration times imply a conscious decision making process. Furthermore, in most of these experiments, only one element on the screen led to a contingent response. In contrast, participants were required to go through a decision making process in the current setup, and we interpreted their decision as an indicator for their preference. However, given the absence of any effects in the side violation phase, it is not clear whether participants actually associated one side of the screen with one type of information, which would be necessary to make an informed decision. Thus, while this paradigm could be useful to study exploratory behaviour and preferences in a controlled setting, its methodological aspects require more in-depth investigations to help us understand the mechanisms participants use in this setting.

The implementation of generalised linear mixed models including the time course of target looking proved fruitful in all three studies. When comparing this statistical approach with the more traditional ANOVA approach, we often found that results across analyses converged or that the GLMM provided additional information. For example, if the time course of target looking only showed an increase towards the end of the trial, ANOVAs and t-tests did not show a significant difference whereas the GLMM was able to capture this response. Although it might not be clear what the underlying process is for such a time course, it suggests that most participants looked to the target during that time. By formulating the theoretical expectations of target looking behaviour, we were able to interpret the data with regard to these theoretical expectations. However, it is still largely unclear how different time courses can be interpreted with regard to participants' processing of information during the learning phase and during the current test trial. Thus, while GLMMs including time appear to be a promising approach for the analysis of child data, more research is needed to understand the theoretical implications of the statistical outcomes.

Theories on word and action learning

As argued in the introduction, single-domain approaches to word and action learning have focused on the respective domain, and therefore, can provide large detail with respect to processes in that domain. However, most of these models are not intended for the integration of two domains or any influences between two domains. Nonetheless, some of them provide an interface that can be used to accomplish the (partial) integration. For example, the PRIMIR model considers context-sensitive variations of phonological realisations (Werker & Curtin, 2005). If we interpret this aspect of the model as context-sensitive word learning, other components in the environment, for example the co-occurrence of actions, could be taken as potential contextual influences on children's word learning. Similarly, any

model that provides an interface with *context* can integrate such cross-domain influences. This example shows that single-domain models can be extended to other domains. Because these models focus on one domain, it is possible to integrate cross-domain influences (such as the one described). However, the implementation is not envisaged, and therefore, can sometimes seem far-fetched. Furthermore, any parallels and differences between processes in different domains cannot be captured in single-domain approaches, and require comparisons between two single-domain theories that each capture one of the two. This implies that such theories cannot predict any behaviour that might be shared across domains, which is crucial to inform future research in this area. Hence, more general approaches to the development of learning seem more suitable for the integration of word and action learning and any cross-domain influences between the two domains.

The Intersensory Redundancy Hypothesis (IRH) makes predictions about the processing of multimodal information. It suggests that, in a multimodal context, stimuli that provide information across senses should facilitate processing, while stimuli with modality-specific information should attenuate processing (Bahrick et al., 2002). For example, Gogate et al. (2006) showed that 6- to 8-month-olds learn syllable-object associations better when their mothers present them with co-occurring actions that are temporally aligned in contrast to asynchronous actions. Based on these results, Gogate et al. (2006) conclude that the temporal synchrony provides redundant information across senses which facilitates word learning in this multimodal context.

In Study 1 and Study 2, we presented words and actions for objects together, albeit temporal synchrony. The limited evidence of word learning below the age of 30 months in our studies extends the study by Gogate et al. (2006) with regard to limited word learning in the presence of asynchronous actions in infants. These findings are in line with the Intersensory Redundancy Hypothesis because modality-

specific information (i.e., auditory words and visual actions) in a multimodal context should attenuate learning.

In contrast to our first two studies, we used temporal synchrony in Study 3 which, based on the IRH, should have facilitated word learning, especially when children are young and processing abilities are still limited. However, 18-month-olds did not learn word-object associations, and for 30-month-olds and 3- to 4-year-olds, word-object association learning performance was dependent on consistency in co-occurring actions across encounters. Thus, while temporal alignment between words and actions might help children learn as suggested by Gogate et al. (2006), their learning success also depends on other factors, such as the form of the input.

While the results from Study 1 and Study 2 are in line with the predictions of the Intersensory Redundancy Hypothesis, we expected children to learn in these studies, especially those older than 24 months since they are likely to be on the other side of the vocabulary spurt, and therefore, experienced word learners already. By presenting both domains separately in Study 1 and Study 2, we wanted to provide children with a context in which it would be easier to process information from both domains. This assumption was based on the results by Sloutsky and Napolitano (2003) who found that 4-year-olds processed both auditory and visual information only when presented separately. However, if children learned in our setting, they showed learning in only one of the two domains (i.e., action-object learning at 12 months and word-object learning at 36 months in Study 1).

Based on their studies, Sloutsky and Napolitano (2003) suggest a trajectory of modality dominance in which children below the age of four show an auditory dominance, children around the age of four appear to be fluctuating between an auditory and a visual dominance, while most adults show a visual dominance (see also Barnhart et al., 2018, Robinson & Sloutsky, 2004). Indeed, 2- to 3-year-olds learned words for objects in Study 1 and 2, and therefore showed an auditory dominance. At 12 months however, children learned actions for objects, and following the argumen-

tation of Robinson and Sloutsky (2004), showed visual dominance. Furthermore, adults learned both words and actions for objects in Study 1, but their target looking in the action-object condition (and not in the word-object condition) was more susceptible to influences such as participants' preferences in Study 2. These results could be taken as evidence that adults showed an auditory dominance while still learning in both modalities. Accordingly, the trajectory of modality dominance appeared to develop from visual to auditory to both domains in the current setting, with a potential advantage of auditory information even in young adulthood (see also Barnhart et al., 2018).

The combination of the IRH and modality dominance therefore generate interesting new hypotheses. While the IRH suggests that temporal alignment should facilitate learning, this might only be true for the dominant domain. According to the studies by Sloutsky and Napolitano (2003), children show an auditory dominance. This could explain why children learned words for objects when temporal alignment was provided in the study by Gogate et al. (2006) since words in this context provide auditory information. In contrast, actions provide visual information which only becomes dominant at later ages (Barnhart et al., 2018), and might therefore be more difficult to learn. Nonetheless, temporal alignment might still support the learner to integrate the action into her current knowledge. Additionally, it could be that the IRH's predictions work well for studies investigating cross-domain influences (such as the influence of temporally aligned actions on word learning) but fail when studies investigate learning for both types of information. Here, the work by Sloutsky and Napolitano (2003) suggests that children learn only one type of information in this setting, which then might be the more dominant one in the context. Based on these predictions, it would be interesting to test whether temporal synchrony between words and actions would facilitate learning also in our setting, and thereby support children to learn in both domains. This would further reflect the interactions between the structure of the input and the learner.

According to evo-devo, interactions are present at any level within the organism, within the environment, and between organism and environment (Lickliter & Honeycutt, 2003). These interactions define development in both phylogenetic and ontogenetic terms. One important aspect throughout development is the acquisition of knowledge that has also been addressed in dynamic systems theories.

Dynamic systems theories assume that the interaction between the learner and the environment is crucial for learning (e.g., Smith & Thelen, 2003). The present set of studies illustrates this interaction: the preferences of the child (Study 2) and the form of the input (Study 3) contribute to learning, and this interaction changes through time with age and the developing abilities of the learner (Study 1, Study 2, and Study 3). We showed that children's learning across the two domains follows different trajectories, and that children's own preferences and the form of the input influenced children's processing of information from the two domains. As such, the present studies provide one concrete example for the interactive nature between the child and the environment with regard to learning: the learner's behaviour changes throughout development, different domains in the environment can influence each other during learning, and the individual characteristics like the preference of the learner and the form of the input interact during learning. As such, the results provide supporting evidence for dynamic systems theories and evo-devo, which assume dynamic interactions to be crucial for development.

As Montessori (1914) and others have noted, the child contributes to her learning experiences by actively engaging with others and exploring the environment. As already discussed above, this was observable in Study 2 when participants' preferences influenced their target looking. Nonetheless, the influence of preferences has its own limits (as seen in the absence of action-object learning in children), which indicates that other factors, such as the environment, can influence children's learning behaviour. Accordingly, this finding already hints towards the interactive nature

between the learner and the environment (see also Smith & Thelen, 2003, Thelen & Smith, 1996).

In Study 3, we showed that the form of the input can influence children's learning behaviour, and as such, this reflects how the environment can influence learning. Also in this setting, we found differences across ages which might be indicative of the contributing role of processing abilities to learning success, which I will further discuss below. Accordingly, this study might in itself reflect the interaction between the state of the learner and the form of the environment.

One assumption of dynamic systems approaches is that variability can induce instability in the learning situation which induces changes in the system, and such changes can be considered learning (Karmiloff-Smith, 1992, Piaget, 1952, Thelen & Smith, 1996, Twomey et al., 2017). Accordingly, one prediction is that variability in the input can help a child to learn (see Twomey et al., 2017). In Study 3, we found that children between 30 and 42 months learned words for objects when provided with consistency, rather than with variability in the input. This contrasts with the prediction of the theory and also with the results observed in the study by Twomey et al. (2017). Potential reasons for these differences, such as the ones outlined above, might be relevant for a more differentiated picture of the influence of variability and consistency with regard to phase shifts and their learning implications.

In all three studies, we tracked the development of the learners' behaviour by presenting different age groups with the same experiments. Across the studies, we see differences in learning behaviour throughout development which suggests that the changing abilities of the learner contribute to differences in learning. This finding is in line with theoretical expectations in dynamic systems approaches that interactions between the learner and the environment change with time given the changing nature of the organism and the environment. As Smith and Thelen (2003) described, these changes can be found on many different time scales. In the current studies, we observed such changes within the experimental procedures when par-

ticipants gained knowledge throughout the learning phases and accessed it in the test phases. Furthermore, we observed such changes across different age groups, and assumed that adults, as a control group, would show the behaviour that should be the norm when processing abilities, learning mechanisms, and knowledge of the world are well developed. Thus, in the present studies it seems that the developing perceptual abilities influenced the learning behaviour. However, we base this assumption on the differentiation of age groups, although it is evident that even within these age groups, learners are at different points in their development and have acquired different kinds of knowledge sets. Accordingly, it would be interesting to investigate which level and which kind of processing abilities a learner requires to integrate the information presented in the current studies.

Because of their wide applicability and adaptability, dynamic systems approaches to development appear to be suited for the theoretical integration of the present studies. Different domains, such as words and actions, their interaction in the input, and the learners' preferences, as well as the interactions between all of these aspects during learning across time can be well integrated in such frameworks to development, and thus, support the idea of dynamic development.

Domain-general vs. domain-specific mechanisms

As outlined above, most of the literature on word and action learning suggests that learning in the two domains reflect certain similarities across the two domains, like the attention to novelty (Dysart et al., 2016). As argued in the introduction, similar behaviours across domains can either be taken as evidence for similar mechanisms (i.e., domain-general), different mechanisms (i.e., domain-specific) that share the temporal trajectory in development, or different mechanisms (i.e., domain-specific) that share a common mediator. In the studies that reveal similarities across domains, the findings are often taken as evidence for domain-general underpinnings across the two domains because the assumed underlying mechanisms (like atten-

tion to novelty) appear to be broad and unspecific to a certain domain (Childers & Tomasello, 2002, 2003, Dysart et al., 2016).

Most of these studies addressed their questions in unimodal contexts by providing children either with novel words or with novel actions for novel objects. In contrast, according to Gogate et al. (2000), caregivers appear to provide their child with both words and actions for an object concurrently. When providing both domains in the same context as in our studies, children seem to learn in one of the two, but not in both domains, suggesting that word and action learning follow different trajectories in early development. Accordingly, learning in the two domains could rely either on domain-specific mechanisms that follow their own temporal schedule, or on domain-general mechanisms that share a common mediator which biases children's learning behaviour.

In the former case, considering domain-specific mechanisms, differing time courses in the development of domain-specific mechanisms would be assumed to be reflected in differing time courses in the development of word and action learning. Such an argumentation would support nativist ideas of development (e.g., Chomsky, 1993) in which it could be assumed that children are born with some form of core knowledge about language and actions. This core knowledge would help children to learn in the two domains, but importantly, the contribution of this knowledge to learning would be specific to the respective domain, and therefore independent across the two domains. Note however, that also a similar developmental learning trajectory across the two domains could be based on domain-specific processes which would then adhere to the same temporal schedule.

In the latter case, considering domain-general mechanisms, it could be that word and action learning rely on domain-general mechanisms as suggested by the studies employing unimodal contexts, but that the two domains share a common mediator that weighs information differently in early development. This approach would support accounts to development that emphasize the role of the environment, and

the interaction between the learner and the environment (e.g., Elman, 2005, Piaget, 1952).

A mediator that weighs information might only surface when a person's processing abilities do not meet the processing demands of the environment. For example, this might be the case when words and actions stand in direct competition for children's attention as it is the case in our study design (see also Sloutsky & Napolitano, 2003). This argumentation would be in line with the predictions of the Intersensory Redundancy Hypothesis as outlined above: if modality-specific information (e.g., visual actions and auditory words) is presented in a multimodal context, learning should be attenuated compared to learning in unimodal contexts. Only if that information is redundant across senses (e.g., temporally aligned), should learning be facilitated. However, the assumption that such a mediator only surfaces when words and actions stand in direct competition is challenged by the findings that 12-month-olds showed action-object learning also in unimodal contexts (Eiteljoerge et al., in prep.b). Instead, it could be that a mediator, such as preference, guides attention to words and actions differently in early development, and thereby influences processing of information, as reflected in Study 2. The influence of the mediator might decrease with any increases in knowledge certainty as discussed in Study 2, which might also relate to processing abilities as studies showing similar processing of words and actions at older ages might indicate (e.g., Dysart et al., 2016). This does not imply the disappearance of the mediator: also adults in Study 2 had individual preferences and their processing was susceptible to such preferences. Only the influence of such a mediator on processing and learning might diminish with increased certainty and processing abilities.

As suggested by Karmiloff-Smith (1992), it could also be that children are born with domain-general mechanisms that are tuned towards domain-specific mechanisms through the environment. With regard to our results, this would suggest that word and action learning in early development are based on shared domain-

general processes (and guided differently by a mediator) but that word and action learning in later development are based on domain-specific processes. This would imply that mechanisms that appear to be shared across domains at later ages (such as attention to novelty in Dysart et al., 2016) are in fact domain-specific. However, assuming different mechanisms for seemingly similar underlying processes seems inefficient, and therefore unlikely. Nonetheless, it could be that certain mechanisms remain domain-general (such as attention to novelty), while others become specific over time if relevant only for a certain domain.

Together, the current studies suggest that the time courses of word and action learning differ across the two domains in early childhood, and therefore render hypotheses that are based on the similarity between these time courses unlikely. However, the studies do not isolate individual mechanisms that are necessary to learn words and actions, and they cannot differentiate further between domain-specific and domain-general processes with regard to children's word and action learning. Accordingly, it would be interesting for future research to investigate the mechanisms that underlie word and action learning in early development at ages below 30 months to identify potential domain-general and domain-specific mechanisms. Such research could help to disentangle the individual contributions of these mechanisms to processing and learning within the two domains, and thereby inform current theories on child development, and more generally, theories on the cognitive structure of human beings.

In addition to the similarities between word and action learning, we investigated potential cross-domain influences between the two domains. More specifically, we investigated how differences in the co-occurrence of actions can influence word learning, and found that consistency in the co-occurrence of actions in contrast to variability in actions across trials helped children learn words for objects in Study 3. These results indicate that information from one domain influenced learning in the other domain. Thus, different domains in the input are not processed separately,

but instead, their co-occurrence provides the learner with additional information that can influence the learning success.

These findings might further be relevant for the interpretations of Study 1 and Study 2: Also here, cross-domain influences might have modulated children's learning success as suggested above when discussing the IRH and modality dominance. That is, children's learning in one domain might have been influenced by the presence of the other domain in the same learning context. To understand the degree of such influences, it would be important to investigate children's learning of both domains while taking cross-domain influences into account. This could be accomplished by a study that combines aspects of Study 1 and Study 3: For example, cross-domain influences from actions to words and from words to actions could be manipulated with regard to the consistency in co-occurrence, and participants could then be tested on both their learning of word-object and action-object associations. Such an approach would provide a more holistic picture of cross-domain influences during learning in a complex environment.

Conclusion

In conclusion, our studies show that children's learning of words and actions for objects transitions from actions to words to both throughout development. In addition, individual preferences of the learner can influence processing of information, and the form of the input can influence the trajectory of learning. Together, these studies address the question of domain-specific and domain-general mechanisms, and reflect the interactive nature of learning between the child and the environment across domains and time.

Other Contributions

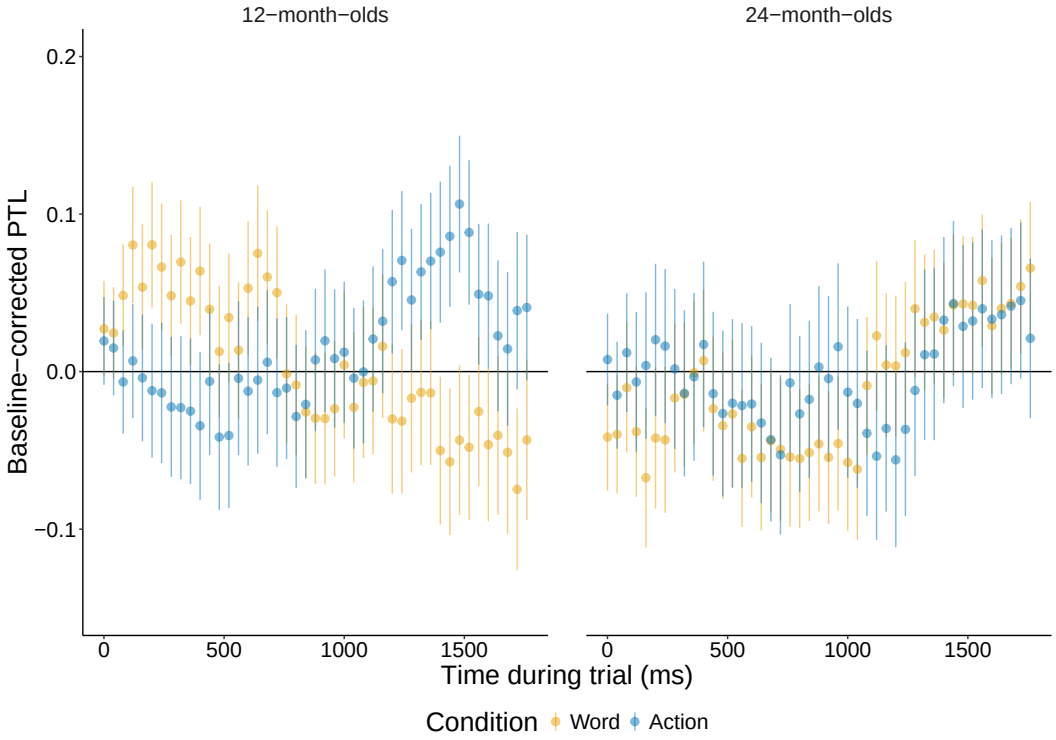
Unimodal word and action learning at 12 and 24 months

Sarah FV Eiteljoerge, Maurits Adam, Birgit Elsner & Nivedita Mani

Studies investigating parallels in word and action learning have mostly studied children between 30 months and 4 years of age and showed that word and action learning seem to develop at a similar time scale and adhere to similar mechanisms such as attention to novelty (Childers & Tomasello, 2002, 2003, Riggs et al., 2015). In most of studies, information was presented unimodally. Thus, children were either presented with a novel word or a novel action for an object. In one study by Hahn and Gershkoff-Stowe (2010) and in the study presented in Chapter the context in which children were presented with words and actions was changed to a bimodal setting where language and actions came together. The latter study showed that although word and action learning in unimodal settings and at later ages seem to develop in synchrony, this picture might not hold in bimodal contexts and with younger children: while 12-month-olds learned action-object but not word-object associations, 36-month-olds learned word-object but not action-object associations. These results suggest that information processing in a bimodal setting might be more challenging than in a unimodal setting (see also Sloutsky & Napolitano, 2003). Thus, we ran the same study as in Chapter 2 but presented the information unimodally to understand how young children's word and action learning develops in unimodal and bimodal contexts.

In this study, we presented 12- and 24-month-old children with words and actions for objects as in Eiteljoerge et al. (2019b) but presented the information in separate eyetracking sessions. Thus, children were presented with two novel objects that were both associated with a novel label and a novel action, and later tested on their learning of these associations. However, in contrast to the previous study, this information was not presented within the same study procedure. Rather, each child

Figure 15. Time course graphs for each of the children's age groups of baseline-corrected proportional target looking (PTL) during the test phase. The word-object condition is represented in yellow, and the action-object condition in blue. The line at 0 represents chance level (increase from baseline), looks above the line indicate target looking and looks below the line indicate distractor looking. The first 240 ms are cut to allow for fixation time, and the time within the trial has been corrected, so that the x-axis starts at 0.



was presented with and tested on one type of association (e.g., word-object association) at the beginning of the test session, and with one type of association (e.g., action-object association) at the end of the test session. Between presentations, children were tested on cognitive and finemotor skills (Bayley, 2006).

The results indicate that 12-month-olds learned action-object associations but not word-object associations, whereas the 24-month-olds learned neither word- nor action-object associations. These results are in line with our previous findings in multimodal contexts and suggest that, independent of uni- and bimodal contexts, children in the current setting might develop from learning action-object to word-object associations in early development.

The impact of verbal cues on 14-and 24-month-olds' visual action anticipation

Maurits Adam, Sarah FV Eiteljoerge, Nivedita Mani & Birgit Elsner

When infants and caregivers interact with each other, a lot of information in the language and the action domain is shared between them. Past research shows that from the first year of life, infants' abilities to process language and action information develop significantly. However, a lot of the research focussed on the language and the action domain in isolation, without investigating in which ways these two domains might inform each other. A growing body of research suggests that verbal information can hinder or facilitate childrens' goal anticipation, but the mechanisms involved are not yet completely understood.

The current study investigated the impact of different verbal information in the form of nouns, verbs, or unspecific information on 14- and 24- month-olds' goal anticipation during observation of a human grasping action. We could not find any evidence for a differential effect of verbal cues on goal anticipation and looking-times in 14- and 24-month-olds. Furthermore, we were not able to find further evidence for the hindering or facilitating effect of verbal cues on action processing.

This experiment was a first step to ensure that verbal cues as well as dependent measures are used in a way that makes the interpretation of results easier. Hence, further fine-grained analyses with comparable use of verbal cues and comparable measures of action anticipation are needed to get a better understanding of the complex mechanisms involved in the potential interactions of the domains of language and action.

References

Adam, M., Eiteljoerge, S., Mani, N., & Elsner, B. (2018). The impact of verbal cues on 14-and 24-month-olds' visual action anticipation. *PsyArXiv. June*, *21*.

Adam, M., Reitenbach, I., & Elsner, B. (2017). Agency cues and 11-month-olds' and adults' anticipation of action goals. *Cognitive Development*, *43*, 37–48. doi: 10.1016/j.cogdev.2017.02.008

Andreu, L., Sanz-Torrent, M., & Trueswell, J. C. (2013). Anticipatory sentence processing in children with specific language impairment: Evidence from eye movements during listening. *Applied Psycholinguistics*, *34*(1), 5–44. doi: 10.1017/s0142716411000592

Ankowski, A. A., Vlach, H. A., & Sandhofer, C. M. (2012). Comparison versus contrast: Task specifics affect category acquisition. *Infant and Child Development*, *22*(1), 1–23. doi: 10.1002/icd.1764

Arias-Trejo, N., & Plunkett, K. (2010). The effects of perceptual similarity and category membership on early word-referent identification. *Journal of Experimental Child Psychology*, *105*(1-2), 63–80. doi: 10.1016/j.jecp.2009.10.002

Arias-Trejo, N., & Plunkett, K. (2013). What's in a link: Associative and taxonomic priming effects in the infant lexicon. *Cognition*, *128*(2), 214–227. doi: 10.1016/j.cognition.2013.03.008

Aslin, R. N., & Shea, S. L. (1990). Velocity thresholds in human infants: Implications for the perception of motion. *Developmental Psychology*, *26*(4), 589. doi: 10.1037/0012-1649.26.4.589

Badets, A., & Osiurak, F. (2015). A goal-based mechanism for delayed motor intention: Considerations from motor skills, tool use and action memory. *Psychological Research*, *79*(3), 345–360. doi: 10.1007/s00426-014-0581-5

Bahrick, L. E., Gogate, L. J., & Ruiz, I. (2002). Attention and memory for faces and actions in infancy: The salience of actions over faces in dynamic events. *Child Development*, *73*(6), 1629–1643. doi: 10.1111/1467-8624.00495

Bahrick, L. E., & Lickliter, R. (2002). Intersensory redundancy guides early perceptual and cognitive development. In R. Kail (Ed.), *Advances in Child Development and Behavior* (Vol. 30, pp. 153–189). New York: Academic Press.

Bahrick, L. E., Lickliter, R., & Flom, R. (2004). Intersensory redundancy guides the development of selective attention, perception, and cognition in infancy. *Current Directions in Psychological Science*, *13*(3), 99–102. doi: 10.1111/j.0963-7214.2004.00283.x

Baldwin, D. A. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language*, *20*(2), 395–418. doi: 10.1017/s0305000900008345

Banks, M., & Salapatek, P. (1978). Acuity and contrast sensitivity in 1-, 2-, and 3-month-old human infants. *Investigative Ophthalmology & Visual Science*, *17*(4), 361–365.

Bard, K. A., & Vauclair, J. (1984). The communicative context of object manipulation in ape and human adult-infant pairs. *Journal of Human Evolution*, *13*(2), 181–190. doi: 10.1016/s0047-2484(84)80062-7

Barnhart, W. R., Rivera, S., & Robinson, C. W. (2018). Different patterns of modality dominance across development. *Acta Psychologica*, *182*, 154–165. doi: 10.1016/j.actpsy.2017.11.017

Baronchelli, A., Chater, N., Pastor-Satorras, R., & Christiansen, M. H. (2012). The biological origin of linguistic diversity. *PloS One*, *7*(10), e48029. doi: 10.1371/journal.pone.0048029

Baron-Cohen, S. (1997). *Mindblindness: An essay on autism and theory of mind.* Cambridge, MA: MIT press.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Bayley, N. (2006). *Bayley scales of infant and toddler development* (3rd ed.). San

Antonio, TX: Psychological Corporation.

Behne, T., Carpenter, M., Call, J., & Tomasello, M. (2005). Unwilling versus unable: infants' understanding of intentional action. *Developmental Psychology*, *41*(2), 328–337. doi: 10.1037/0012-1649.41.2.328

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

Benitez, V. L., & Smith, L. B. (2012). Predictable locations aid early object name learning. *Cognition*, *125*(3), 339–352. doi: 10.1016/j.cognition.2012.08.006

Bentley-Condit, V. K., & Smith, E. (2010). Animal tool use: current definitions and an updated comprehensive catalog. *Behaviour*, 185–221. doi: 10.1163/000579509x12512865686555

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, *109*(9), 3253–3258. doi: 10.1073/pnas.1113380109

Bergelson, E., & Swingley, D. (2013). The acquisition of abstract words by young infants. *Cognition*, *127*(3), 391–397. doi: 10.1016/j.cognition.2013.02.011

Biro, S. (2013). The role of the efficiency of novel actions in infants' goal anticipation. *Journal of Experimental Child Psychology*, *116*(2), 415–427. doi: 10.1016/j.jecp.2012.09.011

Bloom, P. (2002). *How children learn the meanings of words*. Cambridge, MA: MIT press.

Bonawitz, E. B., van Schijndel, T. J., Friel, D., & Schulz, L. (2012). Children balance theories and evidence in exploration, explanation, and learning. *Cognitive Psychology*, *64*(4), 215–234. doi: 10.1016/j.cogpsych.2011.12.002

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and me: Familiar names help launch babies into speech-stream segmentation. *Psychological Science*, *16*(4), 298–304. doi: 10.1111/j.0956-7976.2005.01531.x

Brand, R. J., & Tapscott, S. (2007). Acoustic packaging of action sequences by

infants. *Infancy*, *11*(3), 321–332. doi: 10.1111/j.1532-7078.2007.tb00230.x

Campbell, F. A., Pungello, E. P., Miller-Johnson, S., Burchinal, M., & Ramey, C. T. (2001). The development of cognitive and academic abilities: growth curves from an early childhood educational experiment. *Developmental Psychology*, *37*(2), 231. doi: 10.1037/0012-1649.37.2.231

Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen-through 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behavior and Development*, *21*(2), 315–330. doi: 10.1016/s0163-6383(98)90009-1

Chainay, H., & Humphreys, G. W. (2002). Privileged access to action for objects relative to words. *Psychonomic Bulletin & Review*, *9*(2), 348–355. doi: 10.3758/bf03196292

Cheney, D., Seyfarth, R., & Smuts, B. (1986). Social relationships and social cognition in nonhuman primates. *Science*, *234*(4782), 1361–1366. doi: 10.1126/science.3538419

Cheney, D. L., & Seyfarth, R. M. (1981). Selective forces affecting the predator alarm calls of vervet monkeys. *Behaviour*, *76*(1-2), 25–60. doi: 10.1163/156853981x00022

Childers, J. B., & Tomasello, M. (2002). Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Developmental Psychology*, *38*(6), 967. doi: 10.1037/0012-1649.38.6.967

Childers, J. B., & Tomasello, M. (2003). Children extend both words and nonverbal actions to novel exemplars. *Developmental Science*, *6*(2), 185–190. doi: 10.1111/1467-7687.00270

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Chomsky, N. (1993). *Lectures on government and binding: The pisa lectures* (No. 9). Walter de Gruyter.

Chomsky, N. (2015). *The minimalist program* (20th anniversary ed.). Cambridge, MA: MIT press.

Cook, C., Goodman, N. D., & Schulz, L. E. (2011). Where science starts: Spontaneous experiments in preschoolers' exploratory play. *Cognition*, *120*(3), 341–349. doi: 10.1016/j.cognition.2011.03.003

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, *61*(5), 1584–1595. doi: 10.2307/1130766

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, *208*(4448), 1174–1176. doi: 10.1126/science.7375928

DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior and Development*, *9*(2), 133–155. doi: 10.1016/0163-6383(86)90025-1

Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in Neurosciences*, *21*(8), 355–361. doi: 10.1016/s0166-2236(98)01263-6

Deligianni, F., Senju, A., Gergely, G., & Csibra, G. (2011). Automated gaze-contingent objects elicit orientation following in 8-month-old infants. *Developmental Psychology*, *47*(6), 1499–1503. doi: 10.1037/a0025659

Dysart, E. L. (2018). *Parallels between word and action learning in 3- to 5-year-olds.* (Unpublished doctoral dissertation). University of Hull.

Dysart, E. L., Mather, E., & Riggs, K. J. (2016). Young children's referent selection is guided by novelty for both words and actions. *Journal of Experimental Child Psychology*, *146*, 231–237. doi: 10.1016/j.jecp.2016.01.003

Eiteljoerge, S. F. V., Adam, M., Elsner, B., & Mani, N. (2019a). Consistency of co-occurring actions influences young children's word learning. *Royal Society Open Science*, *6*(8), 190097. doi: 10.1098/rsos.190097

Eiteljoerge, S. F. V., Adam, M., Elsner, B., & Mani, N. (2019b). Word-object and action-object association learning across early development. *PloS One*, *14*(8),

e0220317. doi: 10.1371/journal.pone.0220317

Eiteljoerge, S. F. V., Adam, M., Elsner, B., & Mani, N. (in prep.a). Do preferences for words and actions influence word-object and action-object learning in early childhood?

Eiteljoerge, S. F. V., Adam, M., Elsner, B., & Mani, N. (in prep.b). Unimodal word and action learning at 12 and 24 months.

Elman, J. L. (2005). Connectionist models of cognitive development: where next? *Trends in Cognitive Sciences*, *9*(3), 111–117. doi: 10.1016/j.tics.2005.01.005

Elsner, B., & Pfeifer, C. (2012). Movement or goal: Goal salience and verbal cues affect preschoolers' imitation of action components. *Journal of Experimental Child Psychology*, *112*(3), 283–295. doi: 10.1016/j.jecp.2012.02.010

Fennell, C. T., & Waxman, S. R. (2010). What paradox? Referential Cues Allow for Infant Use of Phonetic Detail in Word Learning. *Child Development*, *81*(5), 1376–1383. doi: 10.1111/j.1467-8624.2010.01479.x

Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT press.

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, *44*(3), 677–694. doi: 10.1017/s0305000916000209

Fulkerson, A. L., & Waxman, S. R. (2007). Words (but not tones) facilitate object categorization: Evidence from 6-and 12-month-olds. *Cognition*, *105*(1), 218–228. doi: 10.1016/j.cognition.2006.09.005

Gampe, A., & Daum, M. M. (2014). Productive verbs facilitate action prediction in toddlers. *Infancy*, *19*(3), 301–325. doi: 10.1111/infa.12047

Gardner, T. J., Naef, F., & Nottebohm, F. (2005). Freedom and rules: the acquisition and reprogramming of a bird's learned song. *Science*, *308*(5724), 1046–1049. doi: 10.1126/science.1108214

Gergely, G., Bekkering, H., & Király, I. (2001). Rational imitation of goal-directed actions in 14-month-olds. In *Proceedings of the Twentythird Annual Confer-*

ence of the Cognitive Science Society (pp. 325–330).

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naıve theory of rational action. *Trends in Cognitive Sciences*, *7*(7), 287–292. doi: 10.1016/s0010-0277(97)00004-8

Gerson, S. A., & Woodward, A. L. (2014). Labels facilitate infants' comparison of action goals. *Journal of Cognition and Development*, *15*(2), 197–212. doi: 10.1080/15248372.2013.777842

Gibson, J. J. (1966). *The senses considered as perceptual systems.* Houghton Mifflin.

Gilbert, S. F., Opitz, J. M., & Raff, R. A. (1996). Resynthesizing evolutionary and developmental biology. *Developmental Biology*, *173*(2), 357–372. doi: 10.1006/dbio.1996.0032

Gliozzi, V., Mayor, J., Hu, J.-F., & Plunkett, K. (2009). Labels as features (not names) for infant categorization: A neurocomputational approach. *Cognitive Science*, *33*(4), 709–738. doi: 10.1111/j.1551-6709.2009.01026.x

Gogate, L. J., & Bahrick, L. E. (1998). Intersensory redundancy facilitates learning of arbitrary relations between vowel sounds and objects in seven-month-old infants. *Journal of Experimental Child Psychology*, *69*(2), 133–149. doi: 10.1006/jecp.1998.2438

Gogate, L. J., Bahrick, L. E., & Watson, J. D. (2000). A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. *Child Development*, *71*(4), 878–894. doi: 10.1111/1467-8624.00197

Gogate, L. J., Bolzani, L. H., & Betancourt, E. A. (2006). Attention to maternal multimodal naming by 6-to 8-month-old infants and learning of word–object relations. *Infancy*, *9*(3), 259–288. doi: 10.1207/s15327078in0903_1

Gogate, L. J., & Hollich, G. (2010). Invariance detection within an interactive system: A perceptual gateway to language development. *Psychological Review*, *117*(2), 496–516. doi: 10.1037/a0019049

Gogate, L. J., & Hollich, G. (2013). Timing matters: Dynamic interactions create sensitive periods for word learning. In *Theoretical and computational models of word learning: Trends in psychology and artificial intelligence* (pp. 28–48). IGI Global. doi: 10.4018/978-1-4666-2973-8.ch002

Gogate, L. J., & Hollich, G. (2016). Early Verb-Action and Noun-Object Mapping Across Sensory Modalities: A Neuro-Developmental View. *Developmental Neuropsychology*, *41*(5-8), 293–307. doi: 10.1080/87565641.2016.1243112

Gogate, L. J., & Maganti, M. (2017). The origins of verb learning: Preverbal and postverbal infants' learning of word–action relations. *Journal of Speech, Language, and Hearing Research*, *60*(12), 3538–3550. doi: 10.1044/2017_jslhr-l-17-0085

Gogate, L. J., Maganti, M., & Bahrick, L. E. (2015). Cross-cultural evidence for multimodal motherese: Asian indian mothers' adaptive use of synchronous words and gestures. *Journal of Experimental Child Psychology*, *129*, 110–126. doi: 10.1016/j.jecp.2014.09.002

Gogate, L. J., Maganti, M., & Laing, K. B. (2013). Maternal naming of object wholes versus parts to preverbal infants: A fine-grained analysis of scaffolding at 6–8 months. *Infant Behavior and Development*, *36*(3), 470–479. doi: 10.1016/j.infbeh.2013.03.012

Gogate, L. J., Prince, C. G., & Matatyaho, D. J. (2009). Two-month-old infants' sensitivity to changes in arbitrary syllable–object pairings: The role of temporal synchrony. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(2), 508–519. doi: 10.1037/a0013623

Gogate, L. J., Walker-Andrews, A. S., & Bahrick, L. E. (2001). The intersensory origins of word-comprehension: an ecological–dynamic systems view. *Developmental Science*, *4*(1), 1–18. doi: 10.1111/1467-7687.00143

Goldenberg, E. R., & Sandhofer, C. M. (2013). Same, varied, or both? Contextual support aids young children in generalizing category labels. *Journal of Exper-*

imental Child Psychology, *115*(1), 150–162. doi: 10.1016/j.jecp.2012.11.011

Gottlieb, J., Oudeyer, P.-Y., Lopes, M., & Baranes, A. (2013). Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends in Cognitive Sciences*, *17*(11), 585–593. doi: 10.1016/j.tics.2013.09.001

Gredebäck, G., & Melinder, A. (2010). Infants' understanding of everyday social interactions: A dual process account. *Cognition*, *114*(2), 197–206. doi: 10.1016/j.cognition.2009.09.004

Grimm, H., Aktaş, M., & Frevert, S. (2010). *SETK 3-5: Sprachentwicklungstest für drei- bis fünfjährige Kinder (3; 0-5; 11 Jahre); Diagnose von Sprachverarbeitungsfähigkeiten und auditiven Gedächtnisleistungen; Manual.* Göttingen: Hogrefe.

Grimm, H., & Doil, H. (2006). *Elternfragebogen für die Früherkennung von Risikokindern. .* Göttingen, Germany: Hogrefe Verlag.

Gutman, L. M., Sameroff, A. J., & Cole, R. (2003). Academic growth curve trajectories from 1st grade to 12th grade: Effects of multiple social risk factors and preschool child factors. *Developmental Psychology*, *39*(4), 777. doi: 10.1037/0012-1649.39.4.777

Hahn, E. R., & Gershkoff-Stowe, L. (2010). Children and adults learn actions for objects more readily than labels. *Language Learning and Development*, *6*(4), 283–308. doi: 10.1080/15475441003635315

He, M., Walle, E. A., & Campos, J. J. (2015). A cross-national investigation of the relationship between infant walking and language development. *Infancy*, *20*(3), 283–305. doi: 10.1111/infa.12071

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

Hepach, R., Vaish, A., & Tomasello, M. (2012). Young children are intrinsically motivated to see others helped. *Psychological Science*, *23*(9), 967–972. doi:

10.1177/0956797612440571

Herbert, J. S. (2011). The effect of language cues on infants' representational flexibility in a deferred imitation task. *Infant Behavior and Development*, *34*(4), 632–635. doi: 10.1016/j.infbeh.2011.06.007

Hirsh-Pasek, K., & Golinkoff, R. M. (1996). The intermodal preferential looking paradigm: A window onto emerging language comprehension. In D. McDaniel, C. McKee, & H. S. Cairns (Eds.), *Language, speech, and communication. methods for assessing children's syntax* (pp. 105–124). Cambridge, Massachusetts: The MIT Press.

Hirsh-Pasek, K., Golinkoff, R. M., & Hollich, G. (2000). An emergentist coalition model for word learning. In *Becoming a word learner: A debate on lexical acquisition* (pp. 136–164). Oxford University Press Oxford. doi: 10.1093/acprof:oso/9780195130324.003.006

Holzen, K. V., & Mani, N. (2012). Language nonselective lexical access in bilingual toddlers. *Journal of Experimental Child Psychology*, *113*(4), 569–586. doi: 10.1016/j.jecp.2012.08.001

Hopstaken, J. F., van der Linden, D., Bakker, A. B., & Kompier, M. A. (2015). The window of my eyes: Task disengagement and mental fatigue covary with pupil dynamics. *Biological psychology*, *110*, 100–106. doi: 10.1016/j.biopsycho.2015.06.013

Horst, J. S. (2013). Context and repetition in word learning. *Frontiers in Psychology*, *4*, 149. doi: 10.3389/fpsyg.2013.00149

Horst, J. S., Parsons, K. L., & Bryan, N. M. (2011). Get the story straight: Contextual repetition promotes word learning from storybooks. *Frontiers in Psychology*, *2*(17). doi: 10.3389/fpsyg.2011.00017

Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, *13*(2), 128–157. doi: 10.1080/15250000701795598

Houston-Price, C., & Nakai, S. (2004). Distinguishing novelty and familiarity effects in infant preference procedures. *Infant and Child Development: An International Journal of Research and Practice*, *13*(4), 341–348. doi: 10.1002/icd.364

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. doi: 10.1037/a0016543

Hunt, G. R. (1996). Manufacture and use of hook-tools by new caledonian crows. *Nature*, *379*(6562), 249. doi: 10.1038/379249a0

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: relation to language input and gender. *Developmental Psychology*, *27*(2), 236. doi: 10.1037/0012-1649.27.2.236

Jackson, I., & Sirois, S. (2009). Infant cognition: going full factorial with pupil dilation. *Developmental Science*, *12*(4), 670–679. doi: 10.1111/j.1467-7687.2008.00805.x

Johnson-Frey, S. H. (2004). The neural bases of complex tool use in humans. *Trends in Cognitive Sciences*, *8*(2), 71–78. doi: 10.1016/j.tics.2003.12.002

Jones, S. S. (2009). The development of imitation in infancy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1528), 2325–2335.

Junge, C., Cutler, A., & Hagoort, P. (2012). Electrophysiological evidence of early word learning. *Neuropsychologia*, *50*(14), 3702–3712. doi: 10.1016/j.neuropsychologia.2012.10.012

Jusczyc, P. (1997). *The discovery of spoken language.* Cambridge, MA: MIT Press.

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, *29*(1), 1–23. doi: 10.1006/cogp.1995.1010

Kanakogi, Y., & Itakura, S. (2011). Developmental correspondence between action prediction and motor ability in early infancy. *Nature Communications*, *2*,

341. doi: 10.1038/ncomms1342

Karasik, L. B., Tamis-LeMonda, C. S., & Adolph, K. E. (2011). Transition from crawling to walking and infants' actions with objects and people. *Child Development*, *82*(4), 1199–1209. doi: 10.1111/j.1467-8624.2011.01595.x

Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science*. Cambridge, MA: MIT Press.

Kenward, B. (2010). 10-month-olds visually anticipate an outcome contingent on their own action. *Infancy*, *15*(4), 337–361. doi: 10.1111/j.1532-7078.2009.00018.x

Kenward, B., Rutz, C., Weir, A. A., & Kacelnik, A. (2006). Development of tool use in new caledonian crows: inherited action patterns and social influences. *Animal Behaviour*, *72*(6), 1329–1343. doi: 10.1016/j.anbehav.2006.04.007

Kidd, C., & Hayden, B. Y. (2015). The psychology and neuroscience of curiosity. *Neuron*, *88*(3), 449–460. doi: 10.1016/j.neuron.2015.09.010

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PloS one*, *7*(5), e36399. doi: 10.1371/journal.pone.0036399

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2014). The goldilocks effect in infant auditory attention. *Child Development*, *85*(5), 1795–1804. doi: 10.1111/cdev.12263

Kirkorian, H. L., Choi, K., & Pempek, T. A. (2016). Toddlers' word learning from contingent and noncontingent video on touch screens. *Child Development*, *87*(2), 405–413. doi: 10.1111/cdev.12508

Koterba, E. A., & Iverson, J. M. (2009). Investigating motionese: The effect of infant-directed action on infants' attention and object exploration. *Infant Behavior and Development*, *32*(4), 437–444. doi: 10.1016/j.infbeh.2009.07.003

Kuhl, P. K. (2007). Is speech learning 'gated' by the social brain? *Developmental science*, *10*(1), 110–120. doi: 10.1111/j.1467-7687.2007.00572.x

Lansink, J. M., & Richards, J. E. (1997). Heart rate and behavioral measures of attention in six-, nine-, and twelve-month-old infants during object exploration. *Child Development*, *68*(4), 610–620. doi: 10.2307/1132113

Lickliter, R., & Bahrick, L. E. (2000). The development of infant intersensory perception: advantages of a comparative convergent-operations approach. *Psychological Bulletin*, *126*(2), 260. doi: 10.1037/0033-2909.126.2.260

Lickliter, R., & Honeycutt, H. (2003). Developmental dynamics: toward a biologically plausible evolutionary psychology. *Psychological Bulletin*, *129*(6), 819–835. doi: 10.1037/0033-2909.129.6.819

Liszkowski, U. (2008). Before l1: A differentiated perspective on infant gestures. *Gesture*, *8*(2), 180–196. doi: 10.1075/gest.8.2.04lis

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin*, *116*(1), 75–98. doi: 10.1037/0033-2909.116.1.75

Mandel, D. R., Jusczyk, P. W., & Pisoni, D. B. (1995). Infants' recognition of the sound patterns of their own names. *Psychological Science*, *6*(5), 314–317. doi: 10.1111/j.1467-9280.1995.tb00517.x

Mani, N., & Ackermann, L. (2018). Why do children learn the words they do? *Child Development Perspectives*. doi: 10.1111/cdep.12295

Mani, N., Johnson, E., McQueen, J. M., & Huettig, F. (2013). How yellow is your banana? toddlers' language-mediated visual search in referent-present tasks. *Developmental Psychology*, *49*(6), 1036–1044. doi: 10.1037/a0029382

Mani, N., & Plunkett, K. (2010). In the infant's mind's ear: Evidence for implicit naming in 18-month-olds. *Psychological Science*, *21*(7), 908–913. doi: 10.1177/0956797610373371

Mani, N., & Plunkett, K. (2011). Phonological priming and cohort effects in toddlers. *Cognition*, *121*(2), 196–206. doi: 10.1016/j.cognition.2011.06.013

Mani, N., & Schneider, S. (2013). Speaker identity supports phonetic category learning. *Journal of Experimental Psychology: Human Perception and Per-*

formance, *39*(3), 623–629. doi: 10.1037/a0030402

Mann, J., Sargeant, B. L., Watson-Capps, J. J., Gibson, Q. A., Heithaus, M. R., Connor, R. C., & Patterson, E. (2008). Why do dolphins carry sponges? *PloS one*, *3*(12), e3868. doi: 10.1371/journal.pone.0003868

Manual, U. (2008). *Tobii x60 and x120 eye-trackers*. Tobii.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. doi: 10.1016/j.jneumeth.2007.03.024

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*(2), 121–157. doi: 10.1016/0010-0285(88)90017-5

Markman, E. M., Wasow, J. L., & Hansen, M. B. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive Psychology*, *47*(3), 241–275. doi: 10.1016/s0010-0285(03)00034-3

Masur, E. F. (1997). Maternal labelling of novel and familiar objects: Implications for children's development of lexical constraints. *Journal of Child Language*, *24*(2), 427–439. doi: 10.1017/s0305000997003115

Matatyaho, D. J., & Gogate, L. J. (2008). Type of maternal object motion during synchronous naming predicts preverbal infants' learning of word–object relations. *Infancy*, *13*(2), 172–184. doi: 10.1080/15250000701795655

Matatyaho-Bullaro, D. J., Gogate, L., Mason, Z., Cadavid, S., & Abdel-Mottaleb, M. (2014). Type of object motion facilitates word mapping by preverbal infants. *Journal of Experimental Child Psychology*, *118*, 27–40. doi: 10.1016/j.jecp.2013.09.010

Mather, E., & Plunkett, K. (2011). Same items, different order: Effects of temporal variability on infant categorization. *Cognition*, *119*(3), 438–447. doi: 10.1016/j.cognition.2011.02.008

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic

and prosodic effects on word segmentation in infants. *Cognitive Psychology*, *38*(4), 465–494. doi: 10.1006/cogp.1999.0721

May, L., Byers-Heinlein, K., Gervain, J., & Werker, J. F. (2011). Language and the newborn brain: does prenatal language experience shape the neonate neural response to speech? *Frontiers in Psychology*, *2*, 222. doi: 10.3389/fpsyg.2011.00222

McMurray, B. (2016). Language at three timescales: The role of real-time processes in language development and evolution. *Topics in Cognitive Science*, *8*(2), 393–407. doi: 10.1111/tops.12201

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*(2), 143–178. doi: 10.1016/0010-0277(88)90035-2

Messer, D. J. (1978). The integration of mothers' referential speech with joint play. *Child Development*, *49*(3), 781–787. doi: 10.2307/1128248

Meyer, M., Hard, B., Brand, R. J., McGarvey, M., & Baldwin, D. A. (2011). Acoustic packaging: Maternal speech and action synchrony. *IEEE Transactions on Autonomous Mental Development*, *3*(2), 154–162. doi: 10.1109/tamd.2010.2103941

Mirman, D. (2014). *Growth curve analysis and visualization using r*. New York: CRC Press. doi: 10.1201/9781315373218

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, *59*(4), 475–494. doi: 10.1016/j.jml.2007.11.006

Montessori, M. (1914). *Dr. montessori's own handbook*. New York, NY: Schocken.

Müller, G. B. (2007). Evo–devo: extending the evolutionary synthesis. *Nature Reviews Genetics*, *8*(12), 943. doi: 10.1038/nrg2219

Namy, L. L., Campbell, A. L., & Tomasello, M. (2004). The changing role of iconic-

ity in non-verbal symbol learning: A u-shaped trajectory in the acquisition of arbitrary gestures. *Journal of Cognition and Development*, *5*(1), 37–57. doi: 10.1207/s15327647jcd0501_3

Namy, L. L., & Waxman, S. R. (1998). Words and gestures: Infants' interpretations of different forms of symbolic reference. *Child Development*, *69*(2), 295–308. doi: 10.2307/1132165

Nicoladis, E., Mayberry, R. I., & Genesee, F. (1999). Gesture and early bilingual development. *Developmental Psychology*, *35*(2), 514–526. doi: 10.1037/0012-1649.35.2.514

Nussenbaum, K., & Amso, D. (2016). An attentional goldilocks effect: An optimal amount of social interactivity promotes word learning from video. *Journal of Cognition and Development*, *17*(1), 30–40. doi: 10.1080/15248372.2015.1034316

Ota, M., & Skarabela, B. (2018). Reduplication facilitates early word segmentation. *Journal of child language*, *45*(1), 204–218. doi: 10.1017/s0305000916000660

Partridge, E., McGovern, M. G., Yung, A., & Kidd, C. (2015). Young children's self-directed information gathering on touchscreens. In *Proceedings of the 37th annual conference of the cognitive science society.*

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. doi: 10.1037/a0023785

Perry, L. K., Samuelson, L. K., Malloy, L. M., & Schiffer, R. N. (2010). Learn locally, think globally: Exemplar variability supports higher-order generalization and word learning. *Psychological Science*, *21*(12), 1894–1902. doi: 10.1177/0956797610389189

Pfeifer, C., & Elsner, B. (2013). Preschoolers' encoding of rational actions: The role of task features and verbal information. *Journal of Experimental Child*

Psychology, *116*(2), 532–544. doi: 10.1016/j.jecp.2012.12.005

Piaget, J. (1936). *Origins of intelligence in the child*. London: Routledge & Kegan Pau.

Piaget, J. (1952). *The origins of intelligence in children* (Vol. 8) (No. 5). New York, NY: International Universities Press.

Puccini, D., & Liszkowski, U. (2012). 15-month-old infants fast map words but not representational gestures of multimodal labels. *Frontiers in Psychology*, *3*, 101. doi: 10.3389/fpsyg.2012.00101

Quinn, P. C., & Bhatt, R. S. (2010). Learning perceptual organization in infancy: The effect of simultaneous versus sequential variability experience. *Perception*, *39*(6), 795–806. doi: 10.1068/p6639

R Core Team. (2016). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Radesky, J. S., & Christakis, D. A. (2016). Keeping children's attention: The problem with bells and whistles. *JAMA pediatrics*, *170*(2), 112–113. doi: 10.1001/jamapediatrics.2015.3877

Rayner, K. (2014). The gaze-contingent moving window in reading: Development and review. *Visual Cognition*, *22*(3-4), 242–258. doi: 10.1080/13506285.2013.879084

Reder, S. M. (1973). On-line monitoring of eye-position signals in contingent and noncontingent paradigms. *Behavior Research Methods & Instrumentation*, *5*(2), 218–228.

Riggs, K. J., Mather, E., Hyde, G., & Simpson, A. (2015). Parallels between action-object mapping and word-object mapping in young children. *Cognitive Science*, *40*(4), 992–1006. doi: 10.1111/cogs.12262

Robinson, C. W., & Sloutsky, V. M. (2004). Auditory dominance and its change in the course of development. *Child Development*, *75*(5), 1387–1401. doi:

10.1111/j.1467-8624.2004.00747.x

Rochat, P. (1989). Object manipulation and exploration in 2-to 5-month-old infants. *Developmental Psychology*, *25*(6), 871. doi: 10.1037/0012-1649.25.6.871

Roembke, T. C., & McMurray, B. (2016). Observational word learning: Beyond propose-but-verify and associative bean counting. *Journal of Memory and Language*, *87*, 105–127. doi: 10.1016/j.jml.2015.09.005

Roseberry, S., Hirsh-Pasek, K., & Golinkoff, R. M. (2014). Skype me! socially contingent interactions help toddlers learn language. *Child Development*, *85*(3), 956–970. doi: 10.1111/cdev.12166

Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, *12*(2), 339–349. doi: 10.1111/j.1467-7687.2008.00786.x

Rost, G. C., & McMurray, B. (2010). Finding the signal by adding noise: The role of noncontrastive phonetic variability in early word learning. *Infancy*, *15*(6), 608–635. doi: 10.1111/j.1532-7078.2010.00033.x

Roy, B. C., Frank, M. C., DeCamp, P., Miller, M., & Roy, D. (2015). Predicting the birth of a spoken word. *Proceedings of the National Academy of Sciences*, *112*(41), 12663–12668. doi: 10.1073/pnas.1419773112

Ruggeri, A., Markant, D., Gureckis, T., & Xu, F. (2016). Active control of study leads to improved recognition memory in children. In *Proceedings of the 38th annual conference of the cognitive science society.*

Russo-Johnson, C., Troseth, G., Duncan, C., & Mesghina, A. (2017). All tapped out: Touchscreen interactivity and young children's word learning. *Frontiers in Psychology*, *8*, 578. doi: 10.3389/fpsyg.2017.00578

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

Samuelson, L. K. (2002). Statistical regularities in vocabulary guide language ac-

quisition in connectionist models and 15-20-month-olds. *Developmental Psychology*, *38*(6), 1016–1037. doi: 10.1037/0012-1649.38.6.1016

Samuelson, L. K., Smith, L. B., Perry, L. K., & Spencer, J. P. (2011). Grounding word learning in space. *PloS One*, *6*(12), e28095. doi: 10.1371/journal.pone.0028095

Satterthwaite, T. D., Green, L., Myerson, J., Parker, J., Ramaratnam, M., & Buckner, R. L. (2007). Dissociable but inter-related systems of cognitive control and reward during decision making: evidence from pupillometry and event-related fmri. *Neuroimage*, *37*(3), 1017–1031. doi: 10.1016/j.neuroimage.2007.04.066

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-prime: User's guide.* Psychology Software Incorporated.

Schwab, J. F., & Lew-Williams, C. (2016). Repetition across successive sentences facilitates young children's word learning. *Developmental Psychology*, *52*(6), 879–886. doi: 10.1037/dev0000125

Sciutti, A., Lohan, K. S., Gredebäck, G., Koch, B., & Rohlfing, K. J. (2016). Language meddles with infants' processing of observed actions. *Frontiers in Robotics and AI*, *3*, 46. doi: 10.3389/frobt.2016.00046

Scofield, J., & Behrend, D. A. (2008). Learning words from reliable and unreliable speakers. *Cognitive Development*, *23*(2), 278–290. doi: 10.1016/j.cogdev.2008.01.003

Seed, A., & Byrne, R. (2010). Animal tool-use. *Current Biology*, *20*(23), R1032–R1039. doi: 10.1016/j.cub.2010.09.042

Seehagen, S., & Herbert, J. S. (2010). The role of demonstrator familiarity and language cues on infant imitation from television. *Infant Behavior and Development*, *33*(2), 168–175. doi: 10.1016/j.infbeh.2009.12.008

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423. doi: 10.1002/j.1538-7305.1948.tb00917.x

Sirois, S., & Jackson, I. R. (2012). Pupil dilation and object permanence in infants.

Infancy, *17*(1), 61–78. doi: 10.1111/j.1532-7078.2011.00096.x

Sloutsky, V. M., & Napolitano, A. C. (2003). Is a picture worth a thousand words? Preference for auditory modality in young children. *Child Development*, *74*(3), 822–833. doi: 10.1111/1467-8624.00570

Smith, L. B. (2004). Shape: A developmental product. In L. Carlson & E. van der Zee (Eds.), *Functional Features in Language and Space: Insights from Perception, Categorization, and Development* (pp. 235–256). London, UK: Oxford University Press. doi: 10.1093/acprof:oso/9780199264339.003.0016

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. doi: 10.1016/j.tics.2018.02.004

Smith, L. B., & Thelen, E. E. (1993). A dynamic systems approach to development: Applications. In *Bradford books series in cognitive psychology.* Cambridge, Massachusetts: The MIT Press.

Smith, L. B., & Thelen, E. E. (2003). Development as a dynamic system. *Trends in Cognitive Sciences*, *7*(8), 343–348. doi: 10.1016/s1364-6613(03)00156-6

Smith, L. B., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568. doi: 10.1016/j.cognition.2007.06.010

Smolak, L., & Weinraub, M. (1983). Maternal speech: strategy or response? *Journal of Child Language*, *10*(2), 369–380. doi: 10.1017/s0305000900007820

Sommerville, J. A., Woodward, A. L., & Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. *Cognition*, *96*(1), B1–B11. doi: 10.1016/j.cognition.2004.07.004

Southgate, V., Chevallier, C., & Csibra, G. (2009). Sensitivity to communicative relevance tells young children what to imitate. *Developmental Science*, *12*(6), 1013–1019. doi: 10.1111/j.1467-7687.2009.00861.x

Stephen, D. G., & Dixon, J. A. (2008). The self-organization of insight: Entropy

and power laws in problem solving. *The Journal of Problem Solving*, *2*(1), 6. doi: 10.7771/1932-6246.1043

Suanda, S. H., Smith, L. B., & Yu, C. (2016). The multisensory nature of verbal discourse in parent–toddler interactions. *Developmental Neuropsychology*, *41*(5-8), 324–341. doi: 10.1080/87565641.2016.1256403

Sullivan, B., Wilson, C. E., & Saldaña, D. (2019). Development of a gaze contingent method for auditory threshold evaluation in non-verbal ASD children. *Research in Autism Spectrum Disorders*, *62*, 85–98. doi: 10.1016/j.rasd.2019.02.006

Sullivan, J. W., & Horowitz, F. D. (1983). Infant intermodal perception and maternal multimodal stimulation: Implications for language development. In C. Rovee-Collier & L. Lipsitt (Eds.), *Advances in infancy research* (pp. 183–239). Norwood, NJ: Ablex: Elsevier.

Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition*, *71*(2), 73–108. doi: 10.1016/s0010-0277(99)00021-9

Szagun, G., Stumper, B., & Schramm, S. A. (2009). *Fragebogen zur frühkindlichen Sprachentwicklung (FRAKIS) und FRAKIS-K (Kurzform).* Frankfurt/Main, Germany: Pearson Assessment.

Taxitari, L., Twomey, K., Westermann, G., & Mani, N. (under review). Only cows and horses: The limits of infants' early word learning.

Terrace, H. S., Son, L. K., & Brannon, E. M. (2003). Serial expertise of rhesus macaques. *Psychological Science*, *14*(1), 66–73. doi: 10.1111/1467-9280.01420

Thelen, E., & Smith, L. B. (1996). *A dynamic systems approach to the development of cognition and action.* Cambridge, Massachusetts: MIT press. doi: 10.1016/s0898-1221(96)90021-8

Tincoff, R., & Jusczyk, P. W. (1999). Some beginnings of word comprehension

in 6-month-olds. *Psychological Science*, *10*(2), 172–175. doi: 10.1111/1467-9280.00127

Tincoff, R., & Jusczyk, P. W. (2011). Six-month-olds comprehend words that refer to parts of the body. *Infancy*, *17*(4), 432–444. doi: 10.1111/j.1532-7078.2011.00084.x

Tincoff, R., Seidl, A., Buckley, L., Wojcik, C., & Cristia, A. (2018). Feeling the way to words: Parents' speech and touch cues highlight word-to-world mappings of body parts. *Language Learning and Development*, *15*(2), 1–23. doi: 10.1080/15475441.2018.1533472

Tomasello, M. (1999). The human adaptation for culture. *Annual Review of Anthropology*, *28*(1), 509–529. doi: 10.1002/9783527619702.ch1

Tomasello, M. (2008). *Origins of human communication*. Cambridge, MA: MIT press.

Torigoe, T. (1985). Comparison of object manipulation among 74 species of non-human primates. *Primates*, *26*(2), 182–194. doi: 10.1007/bf02382017

Twomey, K. E., Ma, L., & Westermann, G. (2017). All the right noises: Background variability helps early word learning. *Cognitive Science*, *42*, 1–26. doi: 10.1111/cogs.12539

Twomey, K. E., Ranson, S. L., & Horst, J. S. (2014). That's more like it: Multiple exemplars facilitate word learning. *Infant and Child Development*, *23*(2), 105–122. doi: 10.1002/icd.1824

Twomey, K. E., & Westermann, G. (2017). Curiosity-based learning in infants: a neurocomputational approach. *Developmental science*, *21*(4), e12629. doi: 10.1111/desc.12629

Van Schaik, C. P., Deaner, R. O., & Merrill, M. Y. (1999). The conditions for tool use in primates: implications for the evolution of material culture. *Journal of Human Evolution*, *36*(6), 719–741. doi: 10.1006/jhev.1999.0304

Vasconcelos, M. (2008). Transitive inference in non-human animals: An empir-

ical and theoretical analysis. *Behavioural Processes*, *78*(3), 313–334. doi: 10.1016/j.beproc.2008.02.017

Vauclair, J. (1996). *Animal cognition: An introduction to modern comparative psychology.* Cambridge, MA: Harvard University Press.

Verbruggen, S. W., Kainz, B., Shelmerdine, S. C., Hajnal, J. V., Rutherford, M. A., Arthurs, O. J., ... Nowlan, N. C. (2018). Stresses and strains on the human fetal skeleton during development. *Journal of the Royal Society Interface*, *15*(138), 20170593. doi: 10.1098/rsif.2017.0593

Vernetti, A., Senju, A., Charman, T., Johnson, M. H., & Gliga, T. (2018). Simulating interaction: Using gaze-contingent eye-tracking to measure the reward value of social signals in toddlers with and without autism. *Developmental Cognitive Neuroscience*, *29*, 21–29. doi: 10.1016/j.dcn.2017.08.004

Vernetti, A., Smith, T. J., & Senju, A. (2017). Gaze-contingent reinforcement learning reveals incentive value of social signals in young children and adults. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1850), 20162747. doi: 10.1098/rspb.2016.2747

Vlach, H. A., & Sandhofer, C. M. (2011). Developmental differences in children's context-dependent word learning. *Journal of Experimental Child Psychology*, *108*(2), 394–401. doi: 10.1016/j.jecp.2010.09.011

Von Frisch, K. (1967). *The dance language and orientation of bees.* Cambridge, MA: Harvard University Press.

Von Holzen, K., Fennell, C. T., & Mani, N. (2019). The impact of cross-language phonological overlap on bilingual and monolingual toddlers' word recognition. *Bilingualism: Language and Cognition*, *22*(3), 476–499. doi: 10.1017/s1366728918000597

Von Holzen, K., & Mani, N. (2012). Language nonselective lexical access in bilingual toddlers. *Journal of Experimental Child Psychology*, *113*(4), 569–586. doi: 10.1016/j.jecp.2012.08.001

Wade, S., & Kidd, C. (2019). The role of prior knowledge and curiosity in learning. *Psychonomic Bulletin & Review*, 1–11. doi: 10.3758/s13423-019-01598-6

Walle, E. A. (2016). Infant social development across the transition from crawling to walking. *Frontiers in Psychology*, *7*, 960. doi: 10.3389/fpsyg.2016.00960

Walle, E. A., & Campos, J. J. (2014). Infant language development is related to the acquisition of walking. *Developmental Psychology*, *50*(2), 336–348. doi: 10.1037/a0033238

Wang, Q., Bolhuis, J., Rothkopf, C. A., Kolling, T., Knopf, M., & Triesch, J. (2012). Infants in control: Rapid anticipation of action outcomes in a gaze-contingent paradigm. *PLoS ONE*, *7*(2), e30884. doi: 10.1371/journal.pone.0030884

Wass, S., Porayska-Pomsta, K., & Johnson, M. H. (2011). Training attentional control in infancy. *Current Biology*, *21*(18), 1543–1547. doi: doi.org/10.1016/j.cub.2011.08.004

Waxman, S. R., Fu, X., Arunachalam, S., Leddon, E., Geraghty, K., & Song, H.-J. (2013). Are nouns learned before verbs? Infants provide insight into a long-standing debate. *Child Development Perspectives*, *7*(3), 155–159. doi: 10.1111/cdep.12032

Waxman, S. R., & Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends in Cognitive Sciences*, *13*(6), 258–263. doi: 10.1016/j.tics.2009.03.006

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word–object associations by 14-month-old infants. *Developmental Psychology*, *34*(6), 1289–1309. doi: 10.1037/0012-1649.34.6.1289

Werker, J. F., & Curtin, S. (2005). Primir: A developmental framework of infant speech processing. *Language Learning and Development*, *1*(2), 197–234. doi: 10.1207/s15473341lld0102_4

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and*

Development, *7*(1), 49–63. doi: 10.1016/s0163-6383(84)80022-3

Westermann, G., & Mani, N. (Eds.). (2017). *Early word learning*. Routledge. doi: 10.4324/9781315730974

Williams, S. E., & Horst, J. S. (2014). Goodnight book: sleep consolidation improves word learning via storybooks. *Frontiers in Psychology*, *5*(184). doi: 10.3389/fpsyg.2014.00184

Yeung, H. H., & Werker, J. F. (2009). Learning words' sounds before learning how words sound: 9-month-olds use distinct objects as cues to categorize speech information. *Cognition*, *113*(2), 234–243. doi: 10.1016/j.cognition.2009.08.010

Yoon, E. Y., Heinke, D., & Humphreys, G. W. (2002). Modelling direct perceptual constraints on action selection: The Naming and Action Model (NAM). *Visual Cognition*, *9*(4-5), 615–661. doi: 10.1080/13506280143000601

Yu, C., & Smith, L. B. (2013). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PLoS ONE*, *8*(11), e79659. doi: 10.1371/journal.pone.0079659

Zack, E., Gerhardstein, P., Meltzoff, A. N., & Barr, R. (2013). 15-month-oldsâĂŹ transfer of learning between touch screen and real-world displays: language cues and cognitive loads. *Scandinavian Journal of Psychology*, *54*(1), 20–25. doi: 10.1111/sjop.12001

Zmyj, N., Buttelmann, D., Carpenter, M., & Daum, M. M. (2010). The reliability of a model influences 14-month-olds' imitation. *Journal of Experimental Child Psychology*, *106*(4), 208–220. doi: 10.1016/j.jecp.2010.03.002

Declaration

Hereby, I declare that all parts of the dissertation were written independently. Assistance of third parties was only accepted if scientifically justifiable and acceptable in regards to the examination regulations, and all sources have been quoted appropriately. Parts of this thesis and some figures have been used in the following articles.

Eiteljoerge, S. F. V., Adam, M., Elsner, B., & Mani, N. (2019a). Consistency of co-occurring actions influences young children's word learning. Royal Society Open Science, 6(8), 190097. doi: 10.1098/rsos.190097

Eiteljoerge, S. F. V., Adam, M., Elsner, B., & Mani, N. (2019b). Word-object and action-object association learning across early development. PloS One, 14 (8), e0220317. doi: 10.1371/journal.pone.0220317

Sarah Eiteljörge

Göttingen, September 2019

Appendix

Study 1: Word-object and action-object association learning across early development

Table 7
Generalized Linear Mixed Models analyzing the effect of condition for each age group to model baseline-corrected Preferential Target Looking over time, including Time and and its linear, quadratic, cubic and quartic term. res <- lmer(PTL_corr.mean ~ Condition(poly1+poly2+poly3+poly4) + Object + Name + Action + WA + (poly1+poly2 || id) + (poly1+poly2 | id:Condition), data=d.adults..gca, REML=F, control = contr)*

Group	Factor	Estimates	*SE*	lower CI	upper CI	LRT	*p*
12 months	Intercept	-0.01	0.03	-0.06	0.04	*	*
	Object	0.01	0.01	-0.01	0.02	0.85	0.36
	Name	0.04	0.01	0.02	0.06	21.13	**<0.001**
	Action	0.04	0.01	0.02	0.06	20.44	**<0.001**
	WA	0.13	0.10	-0.07	0.33	1.60	0.21
	poly1:Condition	-0.26	0.20	-0.63	0.12	1.72	0.19
	poly2:Condition	0.22	0.12	-0.00	0.46	3.30	0.07
	poly3:Condition	0.14	0.06	0.02	0.25	5.58	**0.02**
	poly4:Condition	-0.04	0.06	-0.16	0.07	0.56	0.45
24 months	Intercept	0.01	0.03	-0.05	0.07	*	*
	Object	-0.01	0.01	-0.03	0.01	0.72	0.40
	Name	-0.02	0.01	-0.04	-0.01	6.38	**0.01**
	Action	-0.00	0.01	-0.02	0.02	0.02	0.90
	WA	-0.02	0.10	-0.21	0.18	0.04	0.83
	poly1:Condition	0.22	0.20	-0.16	0.64	1.21	0.27
	poly2:Condition	-0.02	0.13	-0.26	0.27	0.01	0.91
	poly3:Condition	0.02	0.06	-0.10	0.14	0.18	0.67
	poly4:Condition	0.03	0.06	-0.08	0.15	0.29	0.59
36 months	Intercept	0.03	0.02	-0.01	0.07	*	*
	Object	0.03	0.01	0.02	0.05	20.12	**<0.001**
	Name	-0.11	0.01	-0.12	-0.09	206.74	**<0.001**
	Action	0.02	0.01	0.01	0.04	10.55	**<0.001**
	WA	0.03	0.06	-0.10	0.14	0.19	0.66
	poly1:Condition	0.17	0.20	-0.25	0.53	0.68	0.41
	poly2:Condition	-0.21	0.12	-0.44	0.04	2.97	0.08
	poly3:Condition	-0.23	0.05	-0.32	-0.13	20.59	**<0.001**
	poly4:Condition	0.04	0.05	-0.05	0.14	0.77	0.38

Continued on next page

Group	Factor	Estimates	*SE*	lower CI	upper CI	LRT	*p*
Adults	Intercept	0.17	0.05	0.06	0.26	*	*
	Object	-0.07	0.01	-0.08	-0.05	66.96	**<0.001**
	Name	-0.00	0.01	-0.02	0.02	0.02	0.90
	Action	0.03	0.01	0.02	0.05	17.55	**<0.001**
	WA	0.41	0.09	0.22	0.59	14.57	**<0.001**
	poly1:Condition	-0.14	0.15	-0.43	0.15	0.76	0.38
	poly2:Condition	-0.19	0.14	-0.46	0.09	1.78	0.18
	poly3:Condition	0.08	0.05	-0.03	0.18	2.18	0.14
	poly4:Condition	-0.03	0.05	-0.13	0.07	0.43	0.51

*)Note that coefficients of interactions can only be interpreted in relation to the respective baseline levels of the interacting variables. Further, the significance level of intercepts can only be interpreted in a meaningful way when effects on the intercept are tested. Thus, these values are only displayed when they are informative.

Table 8

Generalized Linear Mixed Models for Word and Action conditions for each age group separately to model baseline-corrected Preferential Target Looking over time, including Time and and its linear, quadratic, cubic and quartic term. res <- lmer(PTL_corr.mean ~ (poly1+poly2+poly3+poly4) + Object + Name + Action + WA + (poly1+poly2 || id), data=d.adults.word.gca, REML=F, control = contr)

Group	Factor	Estimates	*SE*	lower CI	upper CI	LRT	*p*
12 months: action	Intercept	-0.05	0.02	-0.10	-0.00	*	*
	poly1	0.14	0.15	-0.14	0.44	0.86	0.35
	poly2	-0.24	0.10	-0.44	-0.03	4.89	**0.03**
	poly3	-0.07	0.04	-0.16	0.00	3.20	0.07
	poly4	0.11	0.04	0.03	0.19	6.61	**0.01**
	Object	0.06	0.01	0.03	0.08	18.29	**<0.001**
	Name	0.06	0.01	0.04	0.09	24.19	**<0.001**
	Action	0.04	0.01	0.01	0.06	9.35	**<0.001**
	WA	-0.07	0.13	-0.33	0.18	0.34	0.56
12 months: words	Intercept	0.02	0.03	-0.04	0.08	*	*
	poly1	-0.14	0.13	-0.40	0.11	1.13	0.29
	poly2	-0.01	0.10	-0.20	0.18	0.01	0.92
	poly3	0.06	0.04	-0.02	0.14	2.18	0.14
	poly4	0.07	0.04	-0.02	0.15	2.63	0.10
	Object	-0.04	0.01	-0.06	-0.02	9.95	**<0.001**
	Name	0.01	0.01	-0.01	0.04	1.19	0.28
	Action	0.04	0.01	0.02	0.06	10.78	**<0.001**
	WA	0.32	0.16	0.01	0.63	3.80	**0.05**
24 months: actions	Intercept	0.04	0.03	-0.02	0.09	*	*
	poly1	-0.29	0.14	-0.56	-0.01	4.11	**0.04**

Continued on next page

Group	Factor	Estimates	*SE*	lower CI	upper CI	LRT	*p*
	poly2	-0.07	0.11	-0.27	0.14	0.43	0.51
	poly3	0.05	0.04	-0.02	0.14	1.72	0.19
	poly4	-0.02	0.04	-0.09	0.06	0.16	0.69
	Object	-0.03	0.01	-0.06	-0.01	6.96	**0.01**
	Name	-0.05	0.01	-0.08	-0.03	15.93	**<0.001**
	Action	-0.00	0.01	-0.03	0.02	0.14	0.70
	WA	-0.06	0.14	-0.32	0.19	0.21	0.65
24 months: words	Intercept	0.02	0.03	-0.05	0.08	*	*
	poly1	-0.06	0.15	-0.34	0.24	0.15	0.70
	poly2	-0.10	0.10	-0.31	0.09	0.99	0.32
	poly3	0.08	0.04	0.00	0.16	4.04	**0.04**
	poly4	0.01	0.04	-0.07	0.09	0.09	0.76
	Object	0.02	0.01	-0.01	0.04	2.15	0.14
	Name	0.00	0.01	-0.02	0.03	0.16	0.69
	Action	0.01	0.01	-0.02	0.03	0.20	0.65
	WA	-0.06	0.16	-0.37	0.24	0.16	0.69
36 months: actions	Intercept	0.01	0.02	-0.02	0.05	*	*
	poly1	-0.15	0.14	-0.42	0.12	1.20	0.27
	poly2	0.11	0.09	-0.06	0.29	1.46	0.23
	poly3	0.14	0.03	0.07	0.20	15.20	**<0.001**
	poly4	-0.04	0.03	-0.11	0.03	1.18	0.28
	Object	0.01	0.01	-0.00	0.03	2.03	0.15
	Name	-0.09	0.01	-0.11	-0.07	78.43	**<0.001**
	Action	0.05	0.01	0.03	0.07	25.41	**<0.001**
	WA	0.03	0.07	-0.11	0.18	0.20	0.66
36 months: words	Intercept	0.15	0.03	0.10	0.20	*	*
	poly1	0.01	0.15	-0.28	0.29	0.01	0.94
	poly2	-0.10	0.09	-0.28	0.07	1.36	0.24
	poly3	-0.09	0.04	-0.16	-0.02	6.68	**0.01**
	poly4	0.01	0.04	-0.06	0.07	0.04	0.85
	Object	0.06	0.01	0.03	0.08	27.53	**<0.001**
	Name	-0.13	0.01	-0.15	-0.11	139.21	**<0.001**
	Action	-0.01	0.01	-0.03	0.01	0.49	0.49
	WA	0.05	0.11	-0.17	0.27	0.20	0.66
Adults: actions	Intercept	0.12	0.05	0.04	0.22	*	*
	poly1	0.12	0.13	-0.14	0.37	0.84	0.36
	poly2	-0.02	0.11	-0.25	0.20	0.04	0.85
	poly3	-0.08	0.04	-0.14	-0.01	4.55	**0.03**
	poly4	-0.04	0.04	-0.11	0.03	1.18	0.28
	Object	0.05	0.01	0.03	0.07	20.80	**<0.001**
	Name	-0.01	0.01	-0.03	0.01	1.37	0.24
	Action	0.04	0.01	0.02	0.06	15.51	**<0.001**

Continued on next page

Group	Factor	Estimates	*SE*	lower CI	upper CI	LRT	*p*
	WA	0.36	0.11	0.16	0.57	9.29	**<0.001**
Adults: words	Intercept	0.22	0.07	0.08	0.36	*	*
	poly1	-0.02	0.10	-0.20	0.17	0.03	0.86
	poly2	-0.20	0.08	-0.37	-0.03	5.54	**0.02**
	poly3	-0.00	0.04	-0.07	0.07	0.01	0.93
	poly4	-0.07	0.04	-0.14	0.00	3.86	**0.05**
	Object	-0.18	0.01	-0.20	-0.16	251.71	**<0.001**
	Name	0.01	0.01	-0.01	0.03	1.14	0.29
	Action	0.02	0.01	0.00	0.05	4.82	**0.03**
	WA	0.47	0.17	0.16	0.81	7.07	**0.01**

*)Note that coefficients of interactions can only be interpreted in relation to the respective baseline levels of the interacting variables. Further, the significance level of intercepts can only be interpreted in a meaningful way when effects on the intercept are tested. Thus, these values are only displayed when they are informative.

Study 2: Do preferences for words and actions influence word-object and action-object learning in early childhood?

Table 9
LMM testing differences between conditions and age groups.
full <- lmer(PTL_corr ~ ConditionSum + age + z.TestAge + (1 | id) + (1 | Object) + (1 | Name) + (1 | Action), data=m.lmm, REML=F, control = contr)

Group	Factor	Estimates	*SE*	*lower CI*	*upper CI*	LRT	*p*
Overall	Intercept	0.24	0.05	0.15	0.32	*	*
	ConditionSum	0.05	0.02	0.02	0.09	8.26	**< 0.001**
	age group	-0.19	0.05	-0.29	-0.10	13.32	**< 0.001**
	z.TestAge	-0.01	0.03	-0.06	0.04	0.08	0.78
Children	Intercept	0.04	0.03	-0.02	0.11	*	0.2
	ConditionSum	0.04	0.02	0.00	0.09	4.40	**0.04**
	z.TestAge	0.02	0.02	-0.02	0.06	0.74	0.39
Adults	Intercept	0.23	0.05	0.13	0.35	*	**< 0.001**
	ConditionSum	0.06	0.03	0.00	0.12	4.11	**0.04**
	z.TestAge	-0.04	0.06	-0.15	0.07	0.64	0.42

*)Note that coefficients of interactions can only be interpreted in relation to the respective baseline levels of the interacting variables. Further, the significance level of intercepts can only be interpreted in a meaningful way when effects on the intercept are tested. Thus, these values are only displayed when they are informative.

Table 10
LMM evaluating the interaction of preference and target looking in the word-object and action-object condition for in children and adults.
pref.words.children.wordsonly <- lmer(PTL_corr ~ Words_F2 + z.TestAge + Words_S2 + (1 | id) + (1 | Object) + (1 | Name) + (1 | Action), data=subset(m.lmm_children, m.lmm_children$Condition=="Word"), REML=F, control = contr)

Group	Factor	Estimates	*SE*	*lower CI*	*upper CI*	LRT	*p*
Children: Words	Intercept	0.09	0.03	0.02	0.15	*	**0.005**
	Words_F2	-0.00	0.04	-0.07	0.08	0.00	0.99
	z.TestAge	0.03	0.03	-0.03	0.09	0.75	0.39
	Words_S2	-0.03	0.04	-0.09	0.05	0.43	0.51
Adults: Words	Intercept	0.30	0.09	0.12	0.50	*	**< 0.001**
	Words_F2	0.04	0.06	-0.08	0.16	0.44	0.51
	z.TestAge	-0.10	0.07	-0.22	0.03	1.94	0.16
	Words_S2	0.10	0.07	-0.04	0.23	2.04	0.15
Children: Actions	Intercept	-0.00	0.06	-0.12	0.11	*	0.96
	Actions_F2	-0.02	0.03	-0.10	0.04	0.50	0.48
	z.TestAge	0.01	0.03	-0.05	0.06	0.08	0.78

Continued on next page

Group	Factor	Estimates	*SE*	*lower CI*	*upper CI*	LRT	*p*
	Actions_S2	0.04	0.03	-0.03	0.10	1.18	0.28
Adults: Actions	Intercept	0.17	0.05	0.07	0.28	*	**0.001**
	Actions_F2	0.08	0.05	-0.02	0.18	2.37	0.12
	z.TestAge	-0.01	0.06	-0.11	0.11	0.02	0.88
	Actions_S2	0.06	0.05	-0.04	0.17	1.37	0.24

*)Note that coefficients of interactions can only be interpreted in relation to the respective baseline levels of the interacting variables. Further, the significance level of intercepts can only be interpreted in a meaningful way when effects on the intercept are tested. Thus, these values are only displayed when they are informative.

Table 11

GLMM testing differences between conditions and age groups over time including Time, its linear, quadratic, and cubic term.

*pref.action.gca.children <- lmer(PTL_corr.mean ~ (poly1+poly2+poly3)*Actions_F2 + z.TestAge + Actions_S2 + (poly1+poly2 | id) + (poly1+poly2 | Object) + (poly1+poly2 | Name) + (poly1+poly2 | Action), data=subset(m.aggr_children, m.aggr_children$Condition=="Action"), REML=F, control = contr)*

Group	Factor	Estimates	*SE*	*lower CI*	*upper CI*	LRT	*p*
Children	Intercept	0.09	0.04	0.001	0.17	*	*
Words	z.TestAge	0.02	0.03	-0.04	0.08	0.33	0.57
	Words_S2	-0.03	0.03	-0.1	0.04	0.59	0.44
	poly1:Words_F2	-0.25	0.16	-0.58	0.08	2.34	0.13
	poly2:Words_F2	-0.05	0.15	-0.33	0.26	0.14	0.71
	poly3:Words_F2	0.1	0.04	0.02	0.19	6	**0.01**
Adults	Intercept	0.31	0.09	0.12	0.49	*	*
Words	z.TestAge	-0.1	0.07	-0.23	0.03	1.68	0.2
	Words_S2	0.11	0.07	-0.03	0.25	2.36	0.13
	poly1:Words_F2	-0.02	0.08	-0.19	0.14	-0.12	1.00
	poly2:Words_F2	< 0.001	0.08	-0.18	0.15	< 0.001	1.00
	poly3:Words_F2	0.03	0.04	-0.05	0.09	0.43	0.51
Children	Intercept	-0.01	0.06	-0.14	0.11	*	*
Actions	z.TestAge	-0.001	0.02	-0.05	0.04	0.002	0.96
	Actions_S2	0.03	0.03	-0.02	0.09	1.18	0.28
	poly1:Actions_F2	-0.05	0.16	-0.37	0.24	0.11	0.74
	poly2:Actions_F2	0.003	0.12	-0.23	0.22	< 0.001	0.98
	poly3:Actions_F2	0.05	0.04	-0.03	0.13	1.62	0.2
Adults	Intercept	0.18	0.07	0.03	0.32	*	*
Actions	z.TestAge	-0.02	0.05	-0.13	0.09	0.13	0.72
	Actions_F2	0.06	0.05	-0.05	0.18	1.27	0.26
	poly1:Actions_F2	0.12	0.09	-0.04	0.3	1.87	0.17

Continued on next page

Group	Factor	Estimates	*SE*	*lower CI*	*upper CI*	LRT	*p*
	poly2:Actions_F2	-0.06	0.1	-0.25	0.13	0.37	0.54
	poly3:Actions_F2	-0.07	0.04	-0.14	-0.01	4.46	**0.03**

*)Note that coefficients of interactions can only be interpreted in relation to the respective baseline levels of the interacting variables. Further, the significance level of intercepts can only be interpreted in a meaningful way when effects on the intercept are tested. Thus, these values are only displayed when they are informative.

Study 3: Consistency of co-occurring actions influences young children's word learning

Table 12
GLMM testing differences between conditions on Proportional Target Looking over time including Time, its linear, quadratic, and cubic term.
*res = lmer(PTL_corr.mean Condition*age*(poly1 + poly2 + poly3) + Object + Label + z.TestAge + (1 + (poly1 + poly2 + poly3) | id), data=d_aggr, REML=F, control = contr)*

Group	Factor	Estimates	*SE*	lower CI	upper CI	LRT	*p*
Overall	Intercept	-0.03	0.05	-0.13	0.06	1)	1)
	Object	0.03	0.00	0.02	0.03	65.29	**< 0.001**
	Label	0.01	0.00	0.00	0.02	7.98	**< 0.001**
	z.TestAge	-0.03	0.11	-0.24	0.18	0.10	0.76
	Condition:age:poly1	1)	1)	1)	1)	8.68	**0.03**
	Condition:age:poly2	1)	1)	1)	1)	9.96	**0.02**
	Condition:age:poly3	1)	1)	1)	1)	3.09	0.38
18	Intercept	-0.09	0.13	-0.35	0.18	1)	1)
	Object	0.04	0.01	0.02	0.05	30.72	**< 0.001**
	Label	0.00	0.01	-0.01	0.01	0.01	0.94
	z.TestAge	-0.22	0.39	-1.02	0.53	0.32	0.57
	Condition:poly1	0.07	0.21	-0.38	0.50	0.09	0.76
	Condition:poly2	0.13	0.13	-0.10	0.37	1.02	0.31
	Condition:poly3	-0.00	0.10	-0.21	0.19	0.00	0.97
30	Intercept	-0.02	0.10	-0.22	0.19	1)	1)
	Object	-0.01	0.01	-0.02	0.00	1.76	0.18
	Label	-0.01	0.01	-0.02	0.00	2.59	0.11
	z.TestAge	0.12	0.23	-0.34	0.60	0.28	0.60
	Condition:poly1	-0.07	0.19	-0.44	0.29	0.15	0.70
	Condition:poly2	-0.27	0.12	-0.48	-0.05	4.99	**0.02**
	Condition:poly3	-0.09	0.08	-0.24	0.07	1.16	0.28
3 - 4	Intercept	0.08	0.13	-0.18	0.36	1)	1)
	Object	0.04	0.01	0.03	0.06	36.99	**< 0.001**
	Label	0.05	0.01	0.04	0.06	48.70	**< 0.001**
	z.TestAge	-0.07	0.10	-0.28	0.12	0.48	0.49
	Condition:poly1	0.50	0.21	0.06	0.87	5.36	**0.02**
	Condition:poly2	0.23	0.14	-0.02	0.53	2.67	0.10
	Condition:poly3	-0.17	0.09	-0.35	-0.01	3.77	**0.05**
Adults	Intercept	0.64	2.15	-3.77	4.90	1)	1)
	Object	0.03	0.01	0.02	0.04	42.12	**< 0.001**
	Label	-0.01	0.01	-0.02	0.00	2.07	0.15
	z.TestAge	0.30	1.52	-2.81	3.31	0.04	0.85

Continued on next page

Group	Factor	Estimates	*SE*	lower CI	upper CI	LRT	*p*
	Condition:poly1	-0.28	0.17	-0.61	0.05	2.80	0.09
	Condition:poly2	0.09	0.10	-0.11	0.29	0.73	0.39
	Condition:poly3	0.01	0.07	-0.13	0.16	0.02	0.88

[1)]Note that coefficients of interactions can only be interpreted in relation to the respective baseline levels of the interacting variables. Further, the significance level of intercepts can only be interpreted in a meaningful way when effects on the intercept are tested. Thus, these values are not displayed here because of limited informativity.

Table 13
GLMM testing successful learning within conditions over time including Time, its linear, quadratic, and cubic term.
res = lmer(PTL_corr.mean (poly1 + poly2 + poly3) + Object + Label + z.TestAge + (1 + (poly1 + poly2 + poly3) | id), data=dadult_Inconsistent, REML=F, control = contr)

Group	Factor	Estimates	*SE*	lower CI	upper CI	LRT	*p*
18	Intercept	0.16	0.27	-0.41	0.70	[1)]	[1)]
Consistent	poly1	0.13	0.15	-0.16	0.41	0.71	0.40
	poly2	0.15	0.09	-0.01	0.34	2.75	0.10
	poly3	0.06	0.07	-0.09	0.20	0.61	0.44
	Object	0.04	0.01	0.02	0.06	12.94	**< 0.001**
	Label	0.04	0.01	0.02	0.06	14.40	**< 0.001**
	z.TestAge	0.68	0.83	-1.03	2.33	0.53	0.47
18	Intercept	-0.09	0.13	-0.36	0.22	[1)]	[1)]
Inconsistent	poly1	0.06	0.15	-0.24	0.37	0.16	0.68
	poly2	0.03	0.08	-0.14	0.20	0.10	0.75
	poly3	0.06	0.07	-0.07	0.19	0.68	0.41
	Object	0.04	0.01	0.02	0.06	17.19	**< 0.001**
	Label	-0.04	0.01	-0.06	-0.02	17.68	**< 0.001**
	z.TestAge	-0.27	0.41	-1.12	0.66	0.42	0.52
30	Intercept	0.17	0.16	-0.16	0.50	[1)]	[1)]
Consistent	poly1	-0.00	0.13	-0.23	0.26	0.00	0.99
	poly2	-0.23	0.08	-0.38	-0.07	6.80	**0.01**
	poly3	-0.04	0.06	-0.16	0.09	0.33	0.56
	Object	-0.02	0.01	-0.03	-0.00	4.12	**0.04**
	Label	-0.04	0.01	-0.05	-0.02	21.04	**< 0.001**
	z.TestAge	-0.17	0.37	-0.94	0.62	0.21	0.65
30	Intercept	-0.11	0.13	-0.35	0.14	[1)]	[1)]
Inconsistent	poly1	0.07	0.14	-0.21	0.32	0.23	0.63
	poly2	0.04	0.08	-0.13	0.20	0.21	0.65
	poly3	0.05	0.05	-0.05	0.16	0.90	0.34
	Object	0.00	0.01	-0.02	0.02	0.01	0.92

Continued on next page

Group	Factor	Estimates	*SE*	lower CI	upper CI	LRT	*p*
	Label	0.02	0.01	-0.00	0.04	3.34	0.07
	z.TestAge	0.29	0.29	-0.32	0.86	0.94	0.33
3 - 4	Intercept	0.31	0.20	-0.13	0.71	[1)]	[1)]
Consistent	poly1	0.36	0.14	0.08	0.61	5.92	**0.01**
	poly2	0.09	0.09	-0.08	0.27	0.98	0.32
	poly3	-0.13	0.07	-0.26	0.00	3.48	0.06
	Object	0.02	0.01	0.00	0.04	5.07	**0.02**
	Label	0.06	0.01	0.04	0.08	37.67	**< 0.001**
	z.TestAge	-0.22	0.15	-0.52	0.11	1.67	0.20
3 - 4	Intercept	-0.07	0.17	-0.44	0.27	[1)]	[1)]
Inconsistent	poly1	-0.14	0.16	-0.45	0.17	0.83	0.36
	poly2	-0.13	0.10	-0.33	0.06	1.69	0.19
	poly3	0.05	0.06	-0.07	0.15	0.64	0.42
	Object	0.07	0.01	0.05	0.09	43.65	**< 0.001**
	Label	0.03	0.01	0.01	0.05	10.48	**< 0.001**
	z.TestAge	0.04	0.13	-0.22	0.32	0.08	0.77
Adults	Intercept	2.49	2.62	-2.91	8.07	[1)]	[1)]
Consistent	poly1	0.16	0.13	-0.09	0.42	1.50	0.22
	poly2	-0.17	0.07	-0.32	-0.03	5.59	**0.02**
	poly3	0.06	0.06	-0.05	0.17	1.28	0.26
	Object	0.04	0.01	0.03	0.06	38.64	**< 0.001**
	Label	-0.00	0.01	-0.02	0.01	0.14	0.70
	z.TestAge	1.64	1.86	-2.20	5.56	0.71	0.40
Adults	Intercept	-6.49	3.55	-14.20	1.22	[1)]	[1)]
Inconsistent	poly1	0.44	0.11	0.22	0.64	13.80	**< 0.001**
	poly2	-0.26	0.08	-0.40	-0.11	9.79	**< 0.001**
	poly3	0.05	0.05	-0.04	0.15	1.17	0.28
	Object	0.02	0.01	0.01	0.04	9.29	**< 0.001**
	Label	-0.01	0.01	-0.02	0.00	1.99	0.16
	z.TestAge	-4.73	2.51	-10.13	0.71	2.76	0.10

[1)]Note that coefficients of interactions can only be interpreted in relation to the respective baseline levels of the interacting variables. Further, the significance level of intercepts can only be interpreted in a meaningful way when effects on the intercept are tested. Thus, these values are not displayed here because of limited informativity.