GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN IN PUBLICA COMMODA SEIT 1737

Are words special in early development?

Ricarda Bothe

Are words special in early development?

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
Eva Ricarda Bothe
aus Langenhagen

Göttingen, April 2023

Betreuungsausschuss

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

Prof. Dr. Birgit Elsner
Entwicklungspsychologie
Universität Potsdam

Prof. Dr. Alexander Gail
Sensorimotor Neuroscience and Neuroprosthetics
Deutsches Primatenzentrum Göttingen

Prüfungskommission

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

Prof. Dr. Birgit Elsner
Entwicklungspsychologie
Universität Potsdam

Weitere Mitglieder der Prüfungskommission

Dr. Tanya Behne
Entwicklungspsychologie
Georg-August-Universität Göttingen

Prof. Dr. Alexander Gail
Sensorimotor Neuroscience and Neuroprosthetics
Deutsches Primatenzentrum Göttingen

Prof. Dr. Anne-Kathrin Schacht
Affective Neuroscience and Psychopysiology
Georg-August-Universität Göttingen

Prof. Dr. Isabell Wartenburger
Patholinguistik and Neurokognition der Sprache
Universität Potsdam

Tag der mündlichen Prüfung: 1. Juni 2023

Acknowledgements

I want thank my supervisor, **Nivi**. Throughout my PhD journey, Nivi's remarkable multitasking skills and unwavering support have been a constant source of inspiration to me. Even when I faced challenges, Nivi's guidance and encouragement helped me overcome my insecurities and stay motivated. I am truly grateful to have had the opportunity to work with and learn from such an incredible mentor and role model. As I embark on the next chapter of my academic journey, I feel especially fortunate to continue working with Nivi. I have no doubt that her guidance and mentorship will be invaluable to me as I pursue my academic goals. Thank you again, Nivi, for inspiring me every step of the way.

I also want to thank **Birgit,** my second supervisor, and acknowledge her contributions to my work. Her willingness to meet with me regularly throughout my doctoral studies was immensely valuable as she provided helpful feedback, especially with regard to research methods. Her support was especially critical during the pandemic when data collection was difficult due to contact restrictions. I cannot thank her enough for making her department's database in Potsdam available to help me overcome these difficulties. I was honored to have the opportunity to give a talk in her department, and I felt incredibly valued as an Early Career Researcher. Thank you, Birgit.

I would like to take this opportunity to thank **Alex Gail**, my third supervisor, for his contributions to my research throughout my PhD journey. Alex's insightful questions and thought-provoking comments during our meetings were instrumental in improving the quality of my work and ensured that I remained protected within the PhD system. Towards the end of my PhD, Alex's guidance and advice proved invaluable in completing my thesis and transitioning into the postdoc stage. I am very grateful to have had Alex as a supervisor, and I look forward to continuing our collaboration in the future. Thank you, Alex, for being an integral part of my academic journey.

I also want to thank **Anne Schacht**, **Tanya Behne**, **Isabel Wartenburger** and **Markus Steinbach** for their support by joining the examination board.

I feel fortunate to have collaborated with my co-authors, **Sarah** and **Léonie**, and our crossing companion, **Maria**. During my PhD, their steady support, exchange of ideas, and mutual encouragement meant the world to me. Together, we navigated the challenges of academic research and grew as individuals and as a team. Thank you, Sarah, Léonie, and Maria, for being such wonderful collaborators and for helping me grow into a better researcher and person.

I would like to thank the **DFG** "Crossing the Borders: The Interplay of Language, Cognition and the Brain in Early Development (FOR 2253), the **RTG2070** "Understanding Social Relationships", and the **Leibniz Science Campus** "Primate Cognition" for their generous financial support and valuable workshops. Their support allowed me to attend international conferences and workshops in San Sebastian, Lancaster, Ottawa, and Vienna, which provided invaluable opportunities for learning and networking.

Roger Mundry had an immeasurable impact on my intellectual growth during my PhD. His guidance on regression models and data analysis, paired with his enthusiasm for the topic, transformed my approach to research. Through his support, Roger made the challenging journey of a PhD more manageable. I am grateful for the knowledge he so generously shared. Thank you, Roger.

I would like to extend my appreciation to **Christina Keller**, our study coordinator, for her dedication, organizational skills, and support throughout my research. Without her contribution, the immense data collection would not have been possible. Thank you, Christina, for your commitment to the success of my research and for always seeking solutions to problems. Your hard work, professionalism, and attention to detail were greatly appreciated, and I am grateful for the opportunity to work with you.

I would like to thank my **colleagues Fatih, Judith, Mariella, Martina, Raji, Sarah, Sebastian and Sia** for their contributions and stimulating discussions throughout this PhD. Your teamwork was invaluable, and I am grateful for the collaborative environment we created that allowed me to seek your advice and insights whenever needed. Your encouragement was critical in helping me overcome challenges and achieve my goals. Thank you for your dedication and for being such an important part of this time.

I would like to thank all the **families** who participated in our studies at WortSchatzInsel. Your valuable contributions were essential to the success of our research, and we could not have achieved our goals without your willingness to participate. I am also grateful for the exceptional support provided by our dedicated **student research assistants**, as well as the 4 **undergraduates** and 3 **graduate students** I got to supervise and who tremendously contributed to the data collection process. Over the past 3 to 4 years, we have conducted numerous testings using eye-tracking in the lab, and coded videos from online experiments. I greatly appreciate your efforts and hard work. I would like to thank Marina Proft and her daughter Nele for kindly allowing us to use photos from our testing sessions for demonstration in this thesis.

I would also like to thank **Rebecca Jürgens** and **Valerie Liebs** for their invaluable support and cooperation during my time in the BeCog graduate school. As a student representative, their guidance and expertise helped me to accurately represent my fellow students and gain a deeper understanding of the graduate academic system. They were always available to answer my questions and offer their help, thank you!

I would like to thank **Anna Fischer** for being my steadfast companion over the last 2 years as we explored the role of mid-level representatives. Anna has not only been a dedicated partner, but we have both fully engaged in the issues that have been discussed in the board meetings. Even during the difficult times that the pandemic brought, I could always count on Anna to be by my side. I am grateful for the insights and knowledge we gained from our discussions. Thank you, Anna, for being such a great partner and friend to me on this journey.

I am extremely grateful to have participated as a mentee in the **Margret Maltby Mentoring** (MMM) program, coordinated by the wonderful **Ulla Heilmeier** and **Frauke Ritter**. Through this program, I gained a sense of security and stability in the academic system and had the privilege of attending valuable workshops on leadership and academic growth for women. Above all, I was fortunate to have **Steffi Höhl** as a mentor who provided invaluable help and support along the way. Steffi committed to regular 1:1 meetings offering a listening ear and sharing her own inspiring story, which helped me tremendously. I am forever grateful for Steffi's time and commitment.

I would like to take this opportunity to also thank the wonderful people in my life who have stood by me on this journey. **Marianne**, our conversations and shared interests have kept me grounded, and I cannot emphasize enough how much I cherish our friendship. This thesis is just one chapter in our story, and I am grateful to have you as a part of it. Your positivity has been instrumental in my success, and I cannot thank you enough for your never-ending support. You have had an invaluable impact on my life and my journey.

Clemens, you have been my most loyal companion for as long as I can remember, and I cannot express enough how much I appreciate your support. Your genuine interest in my growth and development, your love for our bond, and your encouragement to pursue my dreams have been nothing

short of inspiring. I am incredibly fortunate to have you in my life, and I thank you for being a constant source of love and encouragement.

Finally, I would like to express my deepest appreciation to my partner **Alex**. You are my rock, catching me when I stumble, pushing me to achieve my goals, and reminding me of what really matters in life. Your love and support have been unshakable, and I am incredibly grateful for all you have done for me. This achievement is not only mine, but ours as well, because your inspiration and encouragement have brought me to where I am today.

Abstract

In the realm of early development, words have been considered as unique due to their ability to facilitate the formation of physical object categories by highlighting common visual features among them. Nevertheless, infants are frequently exposed to words from an early age, and as such, they are highly familiar with words which can potentially explain word effects in category formation. This thesis aimed to reassess the role of words in early cognition by comparing their effect to that of equally familiar cues from the visual modality, namely actions, which were matched to the arbitrariness of word-object relations. Prior research has shown that infants fail to learn word-object associations in the presence of actions despite the synchronous presentation of these cues in natural interactions with caregivers. To investigate the source of this discrepancy, I administered a study in which infants were presented with words, actions, and objects either synchronously or sequentially. The results indicated that only 2-year-olds exhibited learning of word-object and action-object associations when these cues were presented synchronously, but not sequentially. Conversely, 1-year-olds did not exhibit learning for either cue, suggesting that the demands of multisensory processing in laboratory settings may exceed their processing capacities, thereby disrupting learning of the associations. Although the first study presented in this thesis demonstrated that words and actions are associated with objects in a similar way, we found notable differences in the way that they shape the formation of object categories in the second study. Specifically, we found some evidence that words, but not actions or a combination of words and actions, led infants to form an object category at test, as evidenced by visual novelty preference. Words, therefore, seem to have an advantage in early concept formation, as they allow infants to group physical objects into a single mental representation, a feat that similarly familiar and arbitrary cues from different modalities cannot achieve. In summary, this thesis has highlighted the unique role of words in early development, which contributes to the formation of concepts.

Zusammenfassung

In der frühen Entwicklung gelten Wörter als einzigartig, da sie die Bildung von Objektkategorien erleichtern, indem sie gemeinsame visuelle Merkmale hervorheben. Dennoch sind Kleinkinder mit Wörtern aufgrund der frühen Exposition sehr vertraut, was Wörtern eventuell einen Vorteil in der kognitiven Entwicklung verschaffen könnte. Diese Dissertation untersucht die Rolle von Wörtern im Vergleich zu visuellen Reizen, nämlich Handlungen, die ebenfalls vertraut sind und arbiträr mit Objekten verbunden wurden. Vorangegangene Studien haben gezeigt, dass Kleinkinder keine Wort-Objekt-Assoziationen bilden, wenn gleichzeitig Handlungen präsentiert werden, obwohl diese Kombination in der natürlichen Interaktion mit Betreuungspersonen vorkommt. Die erste Studie in dieser Dissertation wurde daher durchgeführt, um die Gründe für diese Diskrepanz zu untersuchen. Kleinkindern wurden Wörtern, Handlungen und Objekten entweder synchron oder nacheinander präsentiert. Die Ergebnisse zeigten, dass nur zweijährige Kinder Assoziationen zwischen Wörtern und Objekten oder Handlungen und Objekten bilden konnten, wenn diese Reize gleichzeitig präsentiert wurden, nicht aber, wenn sie nacheinander präsentiert wurden. Einjährige Kinder konnten weder Wörter noch Handlungen mit Objekten assoziieren, was darauf hinweist, dass die Anforderungen an die multisensorische Verarbeitung ihre Verarbeitungskapazitäten im Laborkontext übersteigen können. Obwohl die erste Studie gezeigt hat, dass Wörter und Handlungen ähnlich mit Objekten assoziiert werden, gab es Unterschiede in der zweiten Studie bezüglich der Art und Weise, wie diese die Bildung von Objektkategorien beeinflussen. Es wurde beobachtet, dass Wörter, aber nicht Handlungen oder eine Kombination von Wörtern und Handlungen, Kleinkinder dazu brachten eine Objektkategorie zu bilden, wie durch visuelle Neuartigkeitspräferenz gezeigt wurde. Wörter bieten daher einen Vorteil bei der frühen Konzeptbildung, da sie es Kleinkindern ermöglichen, physische Objekte zu einer einzigen mentalen Repräsentation zu gruppieren, was ähnlich vertraute arbiträre Reize aus verschiedenen Modalitäten nicht erreichen können. Insgesamt hebt diese Dissertation die

einzigartige Rolle von Wörtern in der frühen Entwicklung hervor, die zur Bildung von Konzepten beitragen.

Contents

List of Tables

List of Tables Appendix A

List of Tables Appendix B

List of Figures

Chapter 1

General Introduction

Words are believed to play a unique role in early developmental processes because they can direct attention to common perceptual features of objects (e.g., Ferry et al., 2010; Fulkerson & Waxman, 2007). Unlike other auditory cues, words have been shown to override visual details of objects and unify objects with similar names into a common mental representation of their category (e.g., Plunkett et al., 2008). While some argue that this effect is simply due to infants' familiarity with words (e.g., Lewkowicz, 1988; Robinson & Sloutsky, 2004; Sloutsky & Robinson, 2008), others assume that words play a crucial role in initiating category formation processes that go beyond mere naming events. This view is supported by the observation that naming effects persist even when only a limited number of objects in a given category are named (LaTourrette & Waxman, 2019), underscoring the strength and resilience of word cues in shaping category learning. Such findings suggest that words are indeed special and play an essential role in the early development of conceptual organization. Theories based on induction assume that words enable the conceptual organization of visual input by generalizing over single objects (e.g., Lupyan et al., 2007; Lupyan & Bergen, 2016), essentially supervising category formation in a top-down fashion. In contrast, other theories propose that words should be considered as features in the process of category formation and integrated at the level of perceptual features (e.g., Gliozzi et al., 2009; Hu, 2008; Plunkett et al., 2008). In this bottom-up learning approach, infants recognize the perceptual variability of objects with the same name.

In these accounts, word learning is often analyzed in isolation from other social cues, although words are typically accompanied by actions when caregivers demonstrate the use of an object (e.g., Gogate, 2020; Maganti & Gogate, 2013). Such natural interactions involving synchronized word and action cues may provide infants with redundant information that facilitates learning (e.g., Gogate et al., 2001; Gogate & Bahrick, 2001). However, previous laboratory experiments have shown that the simultaneous presentation of word and action cues can impair word learning, possibly due to a lack of synchronization between cues (e.g., Eiteljoerge et al., 2019; Puccini & Liszkowski, 2012). In Chapter 2, I aimed to reexamine the learning of word-action-object pairs in 1- and 2-year-old children, focusing on how the timing of these cues (synchronous versus sequential) affects the acquisition of word-object and action-object pairs. The results showed that 2-year-olds successfully learned word-object and action-object pairs through synchronous presentations, whereas 1-year-olds did not learn under either condition. These findings suggest that word and action cues are similarly associated with objects and highlight the importance of synchronous presentation in improving learning outcomes.

Given the similarities in the impact of words and actions on associative learning, it is possible that these cues also shape object categories in comparable ways. Previous research has shown that goal-directed actions or functions can facilitate category formation both with and without accompanying word cues (Booth & Waxman, 2002). In Chapter 3, we examined the individual and combined contributions of words and actions in a category formation task with 1-year-olds, where actions were matched to the arbitrariness of word-object associations. The results suggest that although word cues can have a positive influence on basic category formation, arbitrary actions or combined word-action cues did not result in a significant learning advantage. While there were some inconsistencies in the analyses, the results suggest that words play a more significant role in early category learning compared to actions. This was observed even when the action cues were specifically designed to match the arbitrary associations between words and objects.

The results from Chapter 3 suggest that word learning and category learning are related in some way, which goes beyond the use of other cues such as actions in category learning. To further investigate the relationship between word learning and category formation, I conduct a third study in which 1-year-olds participate in a word learning task and a category formation task, with half of the infants participating in the word learning task first and the other half participating in the category formation task first. The first task is designed to test whether the infants learn word-object associations or form object categories, and the second task is designed to test the influence of learning from the first task (i.e., word-object knowledge tested for category formation or category knowledge tested for word-object learning). If the two processes are independent, the order of the tasks should not affect infants' learning. However, when they are linked, infants' success on each task depends on their learning on the first task. The study was initially conducted online due to COVID-19 restrictions but then moved back to the laboratory once regulations allowed in-person testing. A brief discussion of the significance of the study and included as a brief report in the dissertation.

In the General Introduction, I will begin by discussing infants' word-object associative learning and word effects in object category learning. Then, I will examine the literature on the processing advantage of words over tones in early development and present theories that explain these effects in terms of top-down and bottom-up processing. However, I will argue that these theories are not sufficient to explain word effects because they ignore the input that infants typically learn from. To determine whether words are indeed special in early development, I will emphasize the importance of comparing them to other cues that are equally frequent and familiar to infants, such as actions that are arbitrarily related to objects. These actions do not serve any functional purpose but simply involve repeatedly moving the objects in a specific way.

Early language learning

Language exposure starts early in life

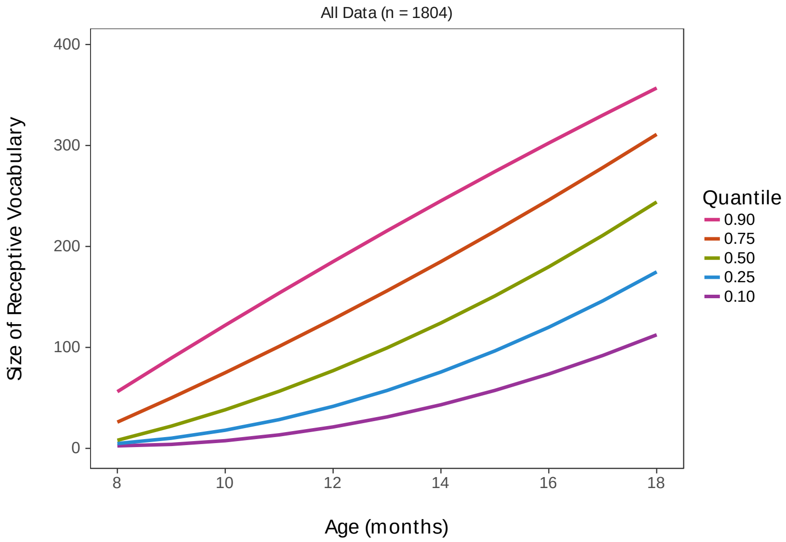
Infant language exposure refers to the amount and variety of language to which an infant is exposed in its environment. By 18 months of age, infants hear between 2,000 and 15,000 spoken words per day (Weisleder & Fernald, 2013), and caregivers who read a single picture book to their infants daily teach them an average of 78,000 novel words per year (Logan et al., 2019). The impact of language exposure has almost immediate consequences for later word learning. For example, the number of words 18-month-olds are exposed to predicts how many words they will know by age 2 (Weisleder & Fernald, 2013). Auditory processing begins early in life, as fetuses respond to sounds between 24 and 27 weeks of gestation (Herschkowitz, 1988; Starr, 1977; Weitzman & Graziani, 1968) and newborns possess rapidly maturing auditory pathways (Lippe et al., 2009), enabling them to recognize voices and express preferences for them. For instance, newborns tend to listen to their mother's voice more frequently than other female voices during high-amplitude sucking experiments (DeCasper & Fifer, 1980), indicating that the ability to recognize familiar voices in their immediate environment may confer evolutionary advantages by establishing a relationship with the group into which they were born and may serve as a foundational block for later language development. Furthermore, fetuses and neonates process auditory cues before visual information (Lippe et al., 2009) and utilize familiar sounds to navigate their environment and form relationships with others as part of social integration (e.g., Kuhl, 2007).

Word-object associative learning

In the early stages of sound processing, infants develop expertise in discriminating sounds in their native language and progressively learn new words to associate with their referents (Werker & Tees, 1984). By the age of one, infants can recognize an average of 50 words for different objects (Tincoff & Jusczyk, 1999) and build a receptive vocabulary of 200 words by 16 months (Fenson et al., 1994), as illustrated in Figure 1. The process of forming associations between words and objects and recognizing objects based on their names starts early in infancy, and repeated experiences strengthen these associations over time (McMurray et al., 2012). Infants frequently encounter the same word-object pairs in their social interactions, which enables them to learn words based on the statistical regularities of these encounters (e.g., Saffran, 2003; Saffran et al., 1996). As a result, repeated exposure to word-object associations leads to more robust recognition of familiar words in their environment, such as "mom," "dad," and "ball" (Bergelson & Swingley, 2012).

Figure 1

Infant's receptive vocabulary from 8 to 18 months



Note. Number of words on the CDI form that an infant growing up in an American-English speaking environment understands (indicated by parent reports) as a function of infant age. Data and graph taken from *The Wordbank Project* (Frank et al., 2016).

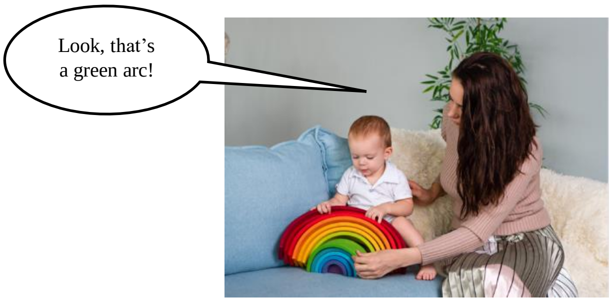
Infants show a remarkable ability to learn novel words quickly, and such word-object associations are more durable than associations between sounds and objects. This finding is supported by the work of MacKenzie et al. (2011), who demonstrated that 12-month-old infants display a preference for words over non-linguistic sounds in an associative learning task. The study employed a switch task paradigm, in which infants were presented with a series of visual pictures of animals, each accompanied by either a word ("fep" or "wug") or a non-linguistic sound (such as "mmmhh", "aaahh", "shhhh"). Following habituation to these associations, infants were presented with both the familiar word-object and sound-object associations (same trials) as well as modified word-object and sound-object associations (switch trials). The results indicated that only infants in the word condition displayed increased looking time for the switch trials as compared to the same trials. In contrast, infants in the sound condition demonstrated similar looking times for both types of trials, indicating that they failed to detect changes in sound-object associations. These findings suggest that infants find it easier to associate words with objects than with sounds, suggesting that words have an advantage in associating with objects relative to other auditory stimuli.

Words as cues to visual category formation

In the context of object naming, infants are frequently exposed to multiple objects that may appear similar. This overlapping distribution of words and objects poses a challenge for infants to determine which object a word refers to (e.g., McMurray et al., 2012). Figure 2 provides an example of this phenomenon in social interactions between infants and their caregivers. Here, the caregiver points or touches a green, arc-shaped object that is part of a group of arc-shaped objects forming a rainbow. This hand movement serves as a cue that helps the infant identify the named object. Consistent naming of similarly shaped objects that differ in color (e.g., a yellow arc, an orange arc, a red arc, etc.) facilitates the recognition of these objects, and eventually enables the infant to form a mental representation of the arc. Through this process, infants can detect patterns in the perceptual variability of objects during the naming process, allowing them to make broader connections between a word and its referent(s). Furthermore, the shape of the object can predict the name of that object, if the name is generalized to other objects of the same shape (referred to as shape bias; Landau et al., 1992). This enables infants to classify objects into basic categories based on their name and perceived similarity.

Figure 2

Caregiver consistently names objects that overlap in perception to facilitate category learning



Thus, words have a significant impact on how infants perceive and remember objects in the context of the perceptual features that indicate category membership (e.g., Althaus & Mareschal, 2014). Moreover, infants form superordinate categories based on a consistent word even when objects look different, e.g., a car and an airplane are both vehicles (Fulkerson & Haaf, 2003). When infants hear a consistent word for objects with large or minimal perceptual overlap, they are able to categorize those objects relative to when they hear no word or a different word for objects (Waxman & Braun, 2005). Evidence for category formation is usually provided by the novelty preference paradigm, in which infants are first habituated to objects that are named the same (e.g., Althaus & Westermann, 2016). They are then presented with two novel objects, one from the familiar category introduced during habituation and an unseen object that looks different from the in-category objects. Visual novelty preference for the out-of-category object is typically interpreted as successful generalization of the in-category objects, including the object from the familiar category at test, which is visually distinguished from the unfamiliar object (for more details on the paradigm, see Methods section).

One of the precursor studies documenting the effect of naming in early category formation was conducted by Waxman & Markow (1995). They used a novelty preference paradigm to examine whether naming objects with the same word facilitated the formation of a basic and a superordinate category for those objects compared to infants who did not hear the word. During object familiarization, 12- to 13-month-old infants were presented with objects from the same perceptual category (e.g., different colored cars) and from different perceptual categories (e.g., different vehicles). Half of the infants saw the objects accompanied by a name in a sentence, e.g., "Look a car" (naming the basic category) or "Look a vehicle" (naming the global category) and the other half with a sentence without a name, "Look what is here." At test, infants saw two novel objects side by side in silence, where one object was another car or vehicle and the other was an object from a different category, such as a fruit or animal. They found that for basic category objects, infants looked at the latter object, which was the target object, longer than they looked at the car, whether or not the objects were presented with a label. This indicates that they formed a basic category for these objects with and without a label. In contrast, only infants in the naming condition looked longer at the object from a different category than the car, indicating generalization of cars and successful formation of a car category. Words, hence, enabled infants to form a superordinate category for cars that they would not otherwise have formed. In the formation of object categories, then, words have a significant role, especially in situations where there is limited perceptual overlap between objects. The following findings suggest that in early development, words are processed more efficiently than sounds because they indicate membership in perceptual categories of objects and can override the perceptual categories of objects.

Processing advantage of words

Words, but not sounds, invite infants to form object categories

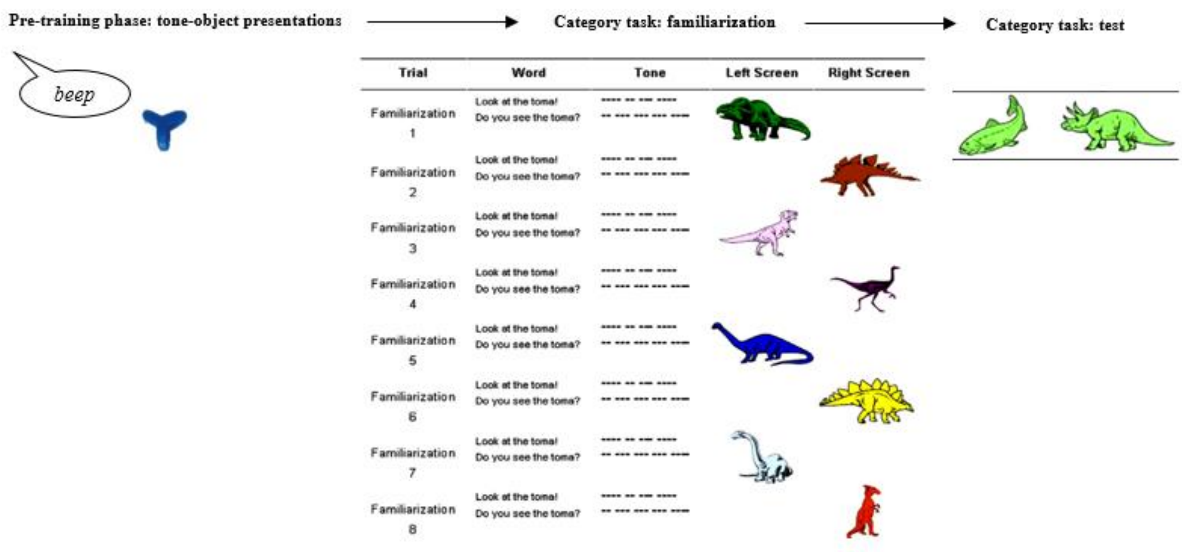
Words were accorded to play a special role in early development because, unlike other auditory stimuli such as sounds, they elicit the formation of exclusive object categories. Such basic categories typically include objects with perceptual overlap and were named the same and exclude objects that do not share perceptual features that are critical for category membership and/or are named differently. For example, Fulkerson & Waxman (2007) familiarized 6- and 12-month-olds with eight different drawings of dinosaurs that differed in shape and color, each accompanied by either the same word in a sentence ("Look at the Toma!") or the same sequence of sounds. At test, all toddlers saw two different drawings of the same color side by side at test, namely a fish and another dinosaur. Both 6- and 12-month-olds in the word condition looked longer at the fish (i.e., introducing a new category) than at the dinosaur from the familiar category. Infants in the sound condition, on the other hand, showed no such preference at test. Thus, novel words presented during object familiarization, but not sounds, may affect how infants perceive the novel, unnamed objects at test and, as here, may lead infants to visually prefer the novel fish to another dinosaur. Thus, naming objects can support the formation of an object category, whereas sounds cannot. Other studies examining the effect of words and sounds in category formation show similar results throughout early development (e.g., Balaban & Waxman, 1997; Ferry et al., 2010; Fulkerson & Haaf, 2003). For example, Ferry and colleagues (2010) used a similar design to the example with dinosaurs described above and showed that the naming effect was surprisingly robust even in infants as young as 4 months of age. In their study, dinosaurs were introduced along with either a consistent word or a consistent sinusoidal tone, and only 4-month-old infants in the naming condition provided evidence for the formation of a dinosaur category on test. Thus, there is ample evidence that words provide a special way of grouping objects that goes beyond the capabilities of simpler auditory stimuli such as tones. Findings such as these have often been taken as evidence that words, in particular, stimulate early cognitive development because they can prompt infants to recognize similarities between objects and group them into mental categories (Booth & Waxman, 2002; Hall & Waxman, 1993; Havy & Waxman, 2016; Lidz et al., 2003; Waxman & Markow, 1995).

Familiar auditory cues tune visual processing

It is possible, however, that the strong effect of naming in early category formation is related to the fact that words are simply more familiar to infants than sounds or tones. Infants are familiar with words because they are common in their environment. Sounds, however, are not among the inputs that infants typically receive and are therefore less familiar than words. In fact, Lany and colleagues (2022) have shown that tones promote object category formation as much as words when infants were familiarized with tone-object pairs prior to the dinosaur category task. In their study, 3-month-old infants initially watched 2-minute videos of novel objects accompanied by a sequence of sounds. For example, in a pre-familiarization trial, infants saw a mill-shaped object bouncing up and down on the screen while they heard a beep (see Figure 3). This pre-habituation with sounds and objects was then followed by the category task frequently used by Waxman, Ferry, and colleagues mentioned above, in which infants saw 8 dinosaurs, one at a time, accompanied by either a beep sequence or a word in a sentence ("Look at the modi"). The objects presented with sounds during pre-familiarization with tones were unrelated to the objects used in the subsequent category formation task. Pre-familiarizing infants with random sound-object pairs before the category learning task showed that the now-familiar sounds promoted categorization at test in a manner similar to words. It is possible, therefore, that exposure to relevant cues (i.e., cues that are subject of the association, such as words or sounds) and the accompanying familiarity with the cue, rather than anything unique about words, drive differences in the extent to which words and unfamiliar sounds trigger the formation of object categories.

Figure 3

Tone-object familiarization in Lany and colleagues (2022) prior to the category formation task designed by Fulkerson & Waxman (2007)



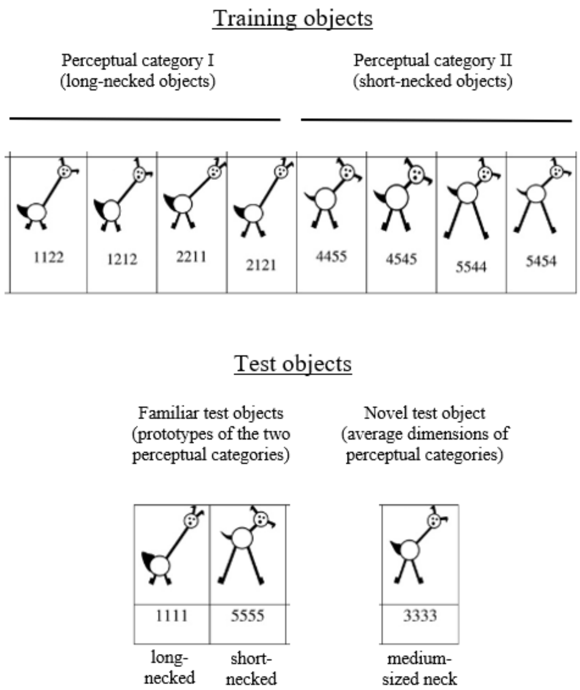
Similar to the results above, Sloutsky and colleagues have shown that 16-month-old infants were able to detect changes in visual cues by familiar sounds (Sloutsky & Robinson, 2008), but not by unfamiliar sounds (Robinson & Sloutsky, 2004). In these studies, one group of infants was familiarized with auditory cues (laser and static tone) without pictures prior to a novelty preference task, so that these infants were familiar with the auditory cues used in the subsequent task, but the pictures of the objects remained novel. Another group of infants was unfamiliar with the auditory cues prior to the actual task, so both the auditory cues and the object pictures were novel (Robinson & Sloutsky, 2004). During the novelty preference procedure, all infants initially saw sound-object pairs, i.e., the image of an object shape (e.g., square, rectangular, or round) was presented along with a sound (the sounds were identical to those presented to infants in Sloutsky & Robinson, 2008 during prefamiliarization). At test, all infants were again presented with the tone-object combinations, with either the tone, the object, or both being novel. Infants who had been familiarized with the tones used in the novelty preference task looked longer at novel objects than at old objects, whereas infants in the other group did not. When infants were familiarized with the tones, they were able to categorize visual objects that they would not otherwise categorize (as did the infants in Robinson & Sloutsky, 2004). Thus, familiar auditory cues supported the processing of visual images, perhaps because familiar stimuli attract more attention than unfamiliar stimuli (Christie & Klein, 1995). That is, familiar auditory cues may have enabled infants to more easily direct their attention to visual stimuli (Sloutsky & Robinson, 2008), which made it easier for them to encode novel objects than when those objects were presented with unfamiliar auditory cues. Thus, familiarity with auditory cues increases attention to visual stimuli, leading to easier processing of visual input. Such tuning effects may be particularly pronounced when visual cues vary widely, for example, when objects look very different and share the same name.

Words override visual object features

Indeed, words can influence the way infants identify objects as common members of a category, even if the objects within that category look different. Evidence for this assumption was provided by Plunkett and colleagues (2008), who showed that the way objects are labeled can influence the way infants generalize these objects and place them into mental categories. Specifically, they first familiarized 10-month-old infants with 8 animated line drawings (i.e., the objects were drawings with heads, legs, tails, and other animated features) that differed in object characteristics such as the length of the neck or legs (see Figure 4). The drawn objects were presented sequentially and infants saw the objects either in silence or in one of 3 labeling conditions. In one of these labeling groups, infants saw the objects together with a label that was always used for objects with a large perceptual overlap. That is, for example, all long-necked objects were named similarly, so that an object with a long neck received a different label than a short-necked object. In a second group, infants saw the objects along with a randomly assigned label so that the label did not correlate with the perceptual overlap of objects given the same name. The third group heard the same label for each object.

Figure 4

Training and test objects used in the category formation task in Plunkett et al. (2008)



Note. Stimulus design and interpretation of looking behavior at test first introduced by Younger (1985).

At test, the same pairs of test objects were presented silently to all infants. One test object was a prototype of one of the 2 categories introduced during familiarization, for example, an object with a long neck (1111) or a short neck (5555). The other test object was an average of these perceptual features (3333), i.e., an object with a medium-sized neck. The latter did not belong to either category presented during the familiarization phase (in which only long-necked and short-necked objects were shown). Plunkett and colleagues found that infants who saw the objects silently or heard two labels that correlated with the objects' perceptual structure visually preferred the medium-necked object that was most unfamiliar to the infants at test, indicating successful category formation. However, when the objects were named randomly, infants showed no signs of category formation at test. Thus, correlated naming promoted object category formation, whereas uncorrelated naming disrupted it. Random labeling was disruptive because infants categorized objects in silence but not when objects were randomly named. Infants at 10 months of age are hence sensitive to the statistical correlations between the labels and the perceptual regularities of the objects. This means that infants soon figure out whether objects that have the same label also look similar, and if they do, they mentally group these objects together. Thus, if a word reflects the structure of visual features, the naming of objects can be 'meaningful' such that the word highlights contrasting visual structures and itself becomes a cue to categorical membership. The latter is also true for infants who were presented with a single label for objects during familiarization, as they treated all objects as members of a single category at test, which they would otherwise divide into two categories (like the infants in the silent condition). Thus, consistently naming objects with little perceptual overlap is sufficient to override the perceptual features of the objects and place them all in a single category, as if the word itself were treated as a marker of category membership. In the consistent naming condition, novelty preference was evidenced by increased attention to one of the prototype objects of the two perceptual categories (1111 & 3333), as these two were indeed "most novel" to the children who formed a single object category. The similarity between the averaged object (5555) and the prototype version that infants hold as a category marker may result in less interest towards the averaged object compared to the prototypes of the two perceptual categories, which may appear more novel and distinct to them.

To sum up, the correlation between words and the perceptual structure of visual input can improve the processing of visual information by enhancing the contrasts in these structures, thereby aiding object categorization. Furthermore, when words are consistently used for objects that differ in their perceptual overlap, they can override these perceptual characteristics in the visual input.

Auditory information overshadows visual information in infancy

The above results on words indicating category membership beyond the perceptual structure of objects are consistent with research showing that auditory input can overshadow visual input in early development (Lewkowicz, 1994; Lewkowicz, 1988). In these studies, 6- and 10-month-old infants were first introduced to two novel stimuli simultaneously, a flashing light and a pulsing tone. Infants were then tested to see if they could recognize these pairs while either the frequency of the visual or auditory stimulus, or both, were changed. Lewkowicz found that infants were more likely to recognize changes in tone frequency than in light frequency. However, when infants were presented with only the flashing light (see Lewkowicz, 1994), they reliably detected changes in these visual stimuli. Thus, infants are perfectly capable of detecting changes in the structure of the flashing light when these cues are presented in isolation, but they looked longer at the combined events when only auditory or auditory and visual cues were changed than if only the visual cues were changed. Thus, when infants have the opportunity to attend to both cues simultaneously (visual and auditory), they are more likely to detect changes in auditory structure. This may imply that auditory cues have a higher level of salience to infants than visual cues, at least in some instances.

Sloutsky & Napolitano (2003) underscored the developmental robustness of such auditory effects when they successfully replicated the infant data for infants with 4-year-old children. Here, children were first presented with visual and auditory combinations, including landscape pictures and simple sounds (such as sine or triangle). During these presentations, children were asked to point to which of these picture-sound combinations indicated the location of a hidden prize, and they were trained to repeatedly select a single picture-sound combination to indicate the location of the prize through feedback reinforcement (i.e., when children pointed to the critical picture, they received a yes, and when they pointed to something else, they received a no). At test, children were again asked where the prize was located while presented with previous combinations of pictures and sounds, but this time either the sound or the picture was changed (new picture + familiar sound or familiar picture + new sound). 4-year-olds relied more on sounds than pictures to indicate the location of the prize, so children were more likely to point to a new picture with a familiar sound than vice versa. This effect remained stable across a range of different visual pictures of varying complexity, so that auditory stimuli dominated visual stimuli even when landscape pictures were particularly salient. Thus, auditory stimuli are not simply more salient than visual stimuli. Why, then, are auditory stimuli so salient? The fact that infants and children rely more on auditory than visual stimuli in such bimodal tasks (auditory and visual information present) could be attributed to more general auditory effects (Lewkowicz, 1994; Lewkowicz, 1988; Robinson & Sloutsky, 2004, 2007, 2008; Sloutsky & Napolitano, 2003; Sloutsky & Robinson, 2008). According to this notion of general auditory effects in early cognitive development, the young brain is particularly sensitive to auditory information when accompanied by visual information, to the extent that processing of visual cues is overshadowed by auditory cues. One possible reason why auditory information produces such overshadowing effects is that auditory information attracts attention more quickly than visual information (e.g., Posner et al., 1976). Increased attention to auditory stimuli relative to visual stimuli could lead to a strong reliance on auditory information in early development and even set the stage for more advanced processing of auditory information in later stages of cognitive and language development (e.g., Robinson & Sloutsky, 2004, 2007, 2008; Sloutsky & Napolitano, 2003; Sloutsky & Robinson, 2008). For example, early auditory dominance may support the development of processes involved in visual similarity recognition and category formation (Sloutsky & Napolitano, 2003).

Words promote categorization even when few objects are labeled

LaTourrette & Waxman (2019) have recently shown that words can trigger visual categorization processes after only a few labeling instances. Using a category formation task as previously described by Waxman and colleagues, they presented 3 groups of 2-year-olds with 8 unfamiliar objects that shared many perceptual features during the familiarization phase. One group saw these objects accompanied by a sentence without a name ("Look over here!"), whereas a second group of children saw the objects accompanied by a sentence that included a consistent name (i.e., each object was accompanied by "Look, it's a modi!"). The third group of infants saw the objects while only the first two objects were named (i.e., early labeling group, "Look, it's a modi!"), while the remaining six objects were referred to with a sentence without a name ("Look over here!"). At test, all children saw two new objects next to each other in silence, a novel member of the category introduced during familiarization and a novel object from an unfamiliar category. In both naming conditions (when either all objects or only the first two objects were named), infants preferred the novel target object to the other object at test, indicating successful formation of the familiar category. Infants who saw the objects without names, however, did not express these visual preferences at test and therefore showed no evidence of category formation. These data illustrate that 2 out of 8 labels are sufficient to trigger category formation at a basic level, similar to the naming of all objects. Infants can thus extend words to other similar-looking objects and then categorize those objects based on the extended label that they would not otherwise have categorized (such as the group that heard no name in the presence of objects). When LaTourrette & Waxman (2019) tested this effect in a fourth group of children in which only the last two objects were labeled with "Look, it's a modi!" (i.e., the late naming labeling group) and the first six objects remained unnamed ("Look over here!"), children showed no evidence of object categorization, similar to the group that had heard no name for any of the objects. Thus, early naming of objects during familiarization promotes category formation at test, whereas late naming does not. While children in the early naming group had the opportunity to extend the new label to other similar-looking objects, children in the late naming group could not extend the new label because no similar-looking objects were presented after naming. That is, 2-year-olds most likely generalize objects based on a name by extending the label to other objects that overlap in perception, but they do not extend labels to mental representations of similar-looking objects within the category that they saw before a label was introduced. Thus, words can initiate object generalization processes when all objects are named or when they are given the opportunity to extend that label to other objects with perceptual overlap that are visually available immediately after naming.

In the previous section, I reviewed studies indicating that words, but not sounds, facilitate category formation of novel objects with limited perceptual overlap, and highlighted how words can influence visual category formation. Recent studies on labeling effects in visual categorization suggest, nonetheless, that words influence category formation processes beyond infants' attention to word-object pairs at the time of experience. Waxman and colleagues discovered that two labels were enough to trigger categorization of multiple objects that received the same name. This suggests that infants might categorize objects based on word extension processes that occur outside of the labeling event, potentially involving additional cognitive factors in categorization process. Consequently, words may be particularly crucial in early cognitive development because they enable infants to establish mental connections between visual objects to the extent that a few instances of labeling initiate categorization processes that continue after the labeling event.

Why might words be special in early cognition?

Induction-based model

Words push for the detection of perceptual commonalities between objects

Studies on the advantage of words over sounds in early category formation suggest that words emphasize common perceptual structures of objects, so that words promote recognition of commonalities among visual objects that share the same name (Althaus & Mareschal, 2014; Althaus & Plunkett, 2015; Booth & Waxman, 2002; Fulkerson & Waxman, 2007; Waxman & Markow, 1995; Waxman & Braun, 2005). Empirical evidence for this assumption was provided by Althaus & Mareschal (2014), who investigated infants' gaze behavior toward novel objects during naming in a category learning task. In their study, 12-month-old infants were first presented with novel objects with perceptual overlap. These objects were presented one at a time, either silently or each object was accompanied by the same word or sound. Similar to the statements by Waxman and colleagues, only the infants in the labeling condition showed signs of category formation at test. When the authors examined the infants' gaze patterns, they found that after only a few word-object presentations, infants in the labeling group (but not the sound group) directed their gaze to object features that they had in common with the previous objects. In the dinosaur example above, this would mean that after looking at a few dinosaurs that all had the same name, infants would be more likely to look at features common to all dinosaurs, such as a dinosaur's tail.

Words, then, may provide conceptual organization of visual input, so that consistent naming of objects leads infants to visually attend to commonalities among those objects. Infants can hence use the information about perceptual similarities highlighted by a consistent word to integrate these similarities into a higher-level mental representation, e.g., they build a dinosaur concept around the representation of individual dinosaurs. Thus, a single word can be associated with many objects so that one word can "supervise" processes that create a shared mental representation of objects named the same (e.g., Booth, 2002; Havy & Waxman, 2016; Waxman & Markow, 1995; Waxman & Braun, 2005). Consistently naming objects with overlapping perceptual features may therefore be the rocky road to naming the category behind them.

Words generalize features and change the way visual cues are seen

But what makes a word a reliable indicator of a category? Lupyan & Bergen (2016) considered the following example: If a monkey and a human are taught to climb a tree to collect coconuts, the monkey must be trained extensively to do so using hand gestures and other pantomimes. The human, however, will understand what to do by simply explaining the task with words, which is more efficient and less time-consuming than explaining with hands. And why? Well, verbal instructions refer more generally to the required climbing movements than simple imitations with the hands (Lupyan & Bergen, 2016). That is, while a word used to describe a climbing action is independent of the fact that each person climbs somewhat differently, an action imitation is specific to the individual action demonstration. Thus, people can use words to generalize individual climbing actions that are irrelevant to the underlying concept of climbing, but still cover all possible climbing actions. A word is, then, a stable reference to a common representation of many individual experiences and thus independent of any single experience.

In addition to the generalizing aspects of words, words can facilitate the recognition of visual features that would otherwise go unnoticed. For example, Lupyan & Ward (2013) have shown that language can influence visual perception in terms of stimulus sensitivity and reaction time. In their study, participants were presented with visual images of familiar objects. However, due to flash suppression, it was very difficult to actually perceive the objects. When participants heard the word *kangaroo* before the presentation of a kangaroo, visual object recognition increased compared to when no label was played, but when they heard something else (that did not match the following image), object recognition decreased. These findings were taken as evidence that labels can change the way objects are recognized, since valid labels (i.e., a label describing an object) enhance object recognition, whereas invalid labels (i.e., a label not referring to an object) suppress it. Thus, it has been described that words have the power to "warp up" the perceptual space because labels can have direct effects on visual perception.

Conceptual knowledge can resolve referential ambiguity

When infants know only few words for objects, it can be challenging to figure out which perceptual commonalities a novel word refers to when objects have different similarities. In particular, in infants' natural environments, there are often objects that share perceptual features with some objects and other perceptual features with other objects. For example, zebras and horses have hooves but tigers do not, and zebras and tigers have stripes but horses do not. So how do children know whether a novel word presented in combination with ambiguous perceptual cues refers to the concept of hooves or stripes? Booth & Waxman (2002) have shown that how children categorize objects based on a novel word may depend on the conceptual context in which these novel word-object pairs occur. In their study, 3-year-old children extended novel labels to other objects based on the shape of the object or on facial features such as eyes, depending on the linguistic context in which the label was presented. During the familiarization phase, all children saw a simple rectangular object with plastic eyes and heard the label embedded in a sentence ("Look it's a dax") in the context of a short story. Half of the children heard the word in a context where the object was made to perform a task, after which the children extended the label to other objects that had the same shape but different eyes. The other half of the children heard the word in a social context associated with words such as "mom" and "love", whereupon the children extended the label to objects that did not have the same shape but had similar eyes. Thus, children use their conceptual knowledge about their mother, who has eyes and often appears in social interactions that include eye contact, to resolve ambiguous naming (when a word might refer to common structures such as eyes or shapes). These results are very striking because animate features (eyes) are particularly salient to young children and therefore could override simpler features such as the shape of an object. However, the children in this study considered linguistic context to infer the commonalities between the perceptual features to which a word refers, and thus resolved ambiguities of novel words. The extent to which novel words alone highlight perceptual commonalities between objects is therefore limited, as children use conceptual knowledge to refine the processes of object generalization and category formation. Familiar words can therefore be used to introduce a concept into a visual scene when concept ambiguity is high.

In summary, words have the ability to draw attention to commonalities between physical objects, leading infants to focus more on perceptual cues indicating category membership when objects are named compared to when they are presented silently or with a sound. Some researchers suggest that words, therefore, warp up the perceptual space which allows for the generalization of objects named in the same way. These processes result in a mental representation of the object category that encompasses all objects sharing perceptual markers of category membership, rather than just a single object. In cases where ambiguity arises in the perception of objects during a naming event, young children rely on their conceptual knowledge gained through familiar words to resolve the ambiguity.

Feature-based model

Words also promote the early formation of object categories when objects look dissimilar. For example, at 9 months of age, infants understand that the word "animal" refers equally to a horse and a giraffe (Fulkerson & Haaf, 2003), and 10-month-old infants combine two distinct perceptual categories when each object is accompanied by the same word (Plunkett et al., 2008). These results suggest that consistent naming of objects leads to more global categories in which perceptual overlap between objects is minimal. In this case, the effect of words in detecting similarities between objects is limited, so infants may instead view the word as a feature of the named object (e.g., Deng & Sloutsky, 2011; Sloutsky & Fisher, 2012). When a consistent word is used to refer to objects that look different, the word itself indicates category membership because the word becomes the commonality between the objects and enables correlations between them (see Younger & Cohen, 1986). Words can therefore have the same status in the formation of categories as other perceptual features that vary in their perceptual salience (Gliozzi et al., 2009). For example, when a category label stands out from other category labels, the contrasting label is recognized as a salient category feature, leading infants to be more likely to categorize objects than when the label is redundant (see also the random labeling condition in Plunkett et al., 2008). In this way, labels do not play a supervisory role in the formation of categories, as suggested by Waxman and colleagues, but instead feed into computations of statistical regularities that infants use to categorize objects and events (Gliozzi et al., 2009; Hu, 2008; Plunkett et al., 2008; Younger & Cohen, 1986). That is, infants are sensitive to the regularities with which certain words and objects occur in their environment, and they use these regularities to make predictions about word-object associations and category membership.

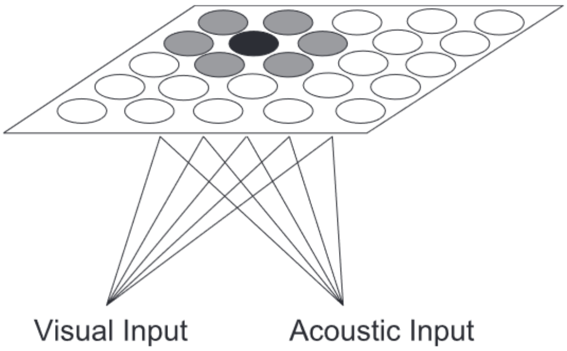
Hu's (2008) study provides evidence for the idea that words can function as features in the process of category learning. Specifically, the study showed that infants were formed categories of basic category objects more easily when presented with accompanying words. However, the study also showed that infants did not actually learn the names of the categories. The study used a structured naming condition, similar to the Plunkett et al. (2008) study, with 10-month-old participants. Infants were shown objects from two perceptual categories, long-necked and short-necked, each accompanied by a name that corresponded to the perceptual category. For example, all long-necked objects were given the same name, but this name was different from the name given to all short-necked objects. At test, infants saw another object from each of the two perceptual categories side by side on the screen (1111 vs. 5555) and heard the name for one of the two categories. Infants did not recognize the test objects by the category labels; that is, they did not direct their gaze at the object when it was labeled, nor did they look longer at the labeled target object than at the other object. Thus, infants can recognize that contrastive labels underly distinct object categories, but they do not learn the word for the concept after only a few category instances. This suggests that the use of labels, rather than the label itself, mediates category formation in early development. Specifically, the results of the above studies show that consistent labeling leads infants to broadly generalize objects with little perceptual overlap, whereas contrastive labeling leads infants to form distinct object categories.

To understand how infants categorize objects without knowing their names, Gliozzi and colleagues (2009) proposed that labels are not simply assigned to objects, but are used as a set of features to distinguish between different categories of objects. They developed a neural network model that integrates both perceptual and labeling features to replicate Plunkett and colleagues' (2008) experiments on infant categorization. In this model, labels are treated equally to other visual markers of category membership, as seen in Figure 5. In this context, both visual and auditory stimuli are included as features in a common mapping and treated equally. In particular, these features are contextualized within a single category of objects. Thus, the network model integrates the modalities of the different features on an equal basis. Empirical findings support the notion that infants use labels as bundles of features to categorize objects. Data from the model's experiments with 10-month-old infants show that infants can classify a new group of objects based on their labels, even though they have different perceptual features. Therefore, labels can be considered as a set of characteristics rather than just names for objects.

Figure 5

Labels as a set of features underlying early categorization (Gliozzi et al., 2009)

V. Gliozzi et al./Cognitive Science 33 (2009)



The architecture: Both visual and acoustic input feed to the same map.

In summary, words play a crucial role in category formation by aiding in the distinction between objects that share commonalities. Infants recognize that different object categories are associated with distinct labels, but they do not immediately learn the word for the concept after a few instances of categorization. Infants may hence view the word as a characteristic of the named object, giving it the same status as perceptual features in category formation. Words, then, can be incorporated into statistical regularity computations that infants use to categorize objects, rather than supervising category formation.

Direction of processing: top-down vs. bottom-up

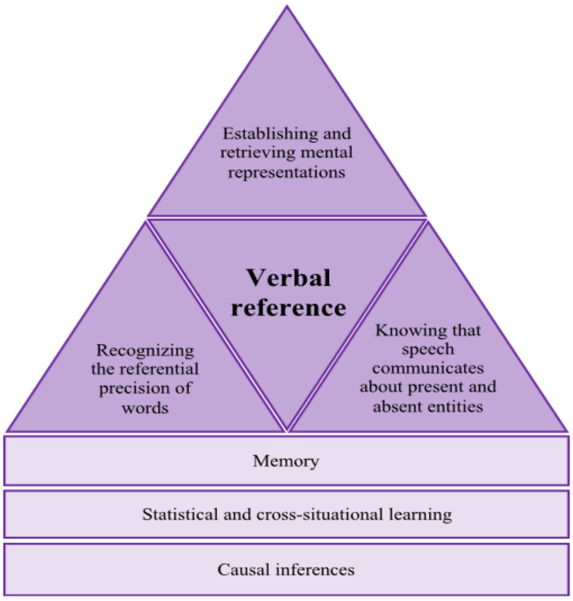
In the feature-based approach to category formation based on words, words are considered as individual features of objects, thus enabling category formation in a bottom-up processing approach (e.g., Gliozzi et al., 2009; Hu, 2008; Plunkett et al., 2008). The bottom-up approach to category learning based on word cues thus emphasizes the significance of sensory input, including auditory and visual stimuli such as object features, and how the brain processes these stimuli. In contrast, induction-based theories assume that words do more than simply representing a feature of an object category, because from this perspective, words change the way objects are perceived (e.g., Lupyan & Bergen, 2016; Lupyan & Ward, 2013). The top-down approach, therefore, emphasizes higher-level cognitive processes such as induction, abstraction, conceptualization, expectations, and the use of context, as demonstrated in the theories proposed by Waxman and colleagues and Lupyan and colleagues. Although these approaches are useful for classifying the theories of feature induction and processing, they may not be sufficient to explain learning because they do not account for the input from which infants learn. In addition to the top-down and bottom-up approaches, an alternative approach to learning proposed by McMurray and colleagues (2012) takes into account the situation at the moment infants learn a novel word and the way learning of that word consolidates over time.

Top-down approaches

Induction-based models use general knowledge or rules to categorize objects when naming objects. These models start from a set of general principles or concepts and apply them to specific objects in order to recognize them. For example, induction-based theories assume that words help infants recognize similarities between objects because words introduce a system into the visual scene (e.g., Booth & Waxman, 2002; Havy & Waxman, 2016; Luchkina & Waxman, 2021; Lupyan & Bergen, 2016; Waxman & Markow, 1995; Waxman & Braun, 2005). Luchkina and Waxman (2021) argue that the comprehension of referential knowledge involves understanding that a word refers to a specific object or group of objects. As illustrated in Figure 6, knowledge about what a word refers to relies on fundamental cognitive abilities such as attention, memory, and categorization. Such cognitive abilities are supported by a range of linguistic reference skills. To acquire this knowledge, children must possess a detailed and nuanced understanding of how language references objects, enabling them to form word-object associations, retrieve them from memory, differentiate between accurate and ambiguous associations, and recognize that referred objects may not be in their immediate surroundings. Waxman and colleagues have developed a theoretical framework that unifies these factors into a model of verbal reference which has the ability to regulate attention in both statistical and cross-situational learning environments. This process is critical for children to recognize important features of objects and events in the world, retain information about these objects in memory, and form conceptual representations of objects and events that can be linked to words. Hierarchical models therefore suggest that top-down mechanisms are the primary catalysts for learning and lead infants to form categories when confronted with named objects.

Figure 6

Linguistic abilities are hierarchically superior to more general processing mechanisms (Luchkina & Waxman, 2021)



Bottom-up approaches

Feature-based models recognize objects based on their perceptual features or properties. These models decompose objects into their constituent parts or features and then use these features to recognize them. Categories are thus defined by the presence or absence of certain features, and objects that share similar features are grouped together. At its core, feature-based categorization is a bottom-up approach because it emphasizes the importance of perceptual information in the categorization process. According to this approach, categories are not pre-existing concepts that we apply to the world but arise from our observations of the world around us (e.g., Smith, 2005). By paying attention to the features of objects, infants can categorize them based on their shared perceptual properties. Feature-based categorization is therefore flexible because categories are defined based on perceptual features rather than on pre-existing concepts.

Situational and slow associative learning

Such induction- and feature-based explanations have been challenged, however, because the association of a word with an object, or the formation of a common category of objects when they share the same name, is not standardized in learning. This means that word-object associations are not always similar in form and relationship (McMurray et al., 2012), making some word-object associations easier to learn than others. McMurray and colleagues (2012) instead proposed a model of word learning that focuses on the moment in which learning occurs, and how learning unfolds over time. Their study used computer simulations of situational processing and slow associative learning and showed that the proposed model can explain a wide range of empirical data on word learning, including the learning of novel words and the acquisition of word meanings over time. Below I will briefly summarize the mechanisms behind situational processing and slow associative learning and argue that these two approaches explain learning processes more holistically because they take into account the context in which the novel word-object pairs are associated.

First, online referent selection, i.e., when infants select a referent for a novel word in the moment, is responsible for the rapid mapping of novel words to their referents. Such “fast mapping” is a mechanism by which infants quickly form an initial association between a novel word and its referent based on a single exposure or a few exposures. This mechanism enables infants to learn the meaning of novel words in real time, even in complex and unpredictable situations in which children must quickly infer the meaning of a novel word from contextual cues or other sources of information. Second, slow associative learning is responsible for the gradual acquisition of word meanings over time. This mechanism allows children to learn statistical regularities in the linguistic environment, such as the co-occurrence of words and their references, and to use this information to expand their vocabulary over time. Slow associative learning, then, involves accumulating evidence over multiple encounters with a word and its referent so that infants can build a more robust and accurate idea of the word's meaning.

In this way, young children can learn novel associations and refine their understanding of them over time. These approaches succeed in addressing a much larger theoretical problem, namely, how infants learn novel words for objects or novel categories based on a word. Unlike induction- and feature-based models, situational processing and slow associative learning models take into account the context of learning in the moment and over time, such as the infant's learning environment with social interactions that provide the input from which infants learn. Instead, infants can learn word-object associations and object categories when objects are named based on the input they receive in the learning context. In the following section, I show that word learning does not occur in isolation, suggesting that input drives associative and category learning in naming objects in early development.

The role of input in early learning

Words appear mostly in multimodal contexts

Indeed, the social interactions in which infants typically learn the meaning of novel words involve information from a variety of domains, such as auditory and visual. Word learning is therefore naturally embedded in interactions with a partner who provides infants with multimodal information. Indeed, caregivers actively adapt this multimodal input to their infants' perceptual abilities during communication to make the informative cues easily accessible by modifying their speech (Golinkoff et al., 2015; Thiessen et al., 2010), use of gestures (Brand et al., 2002; Gogate et al., 2000; Goldin-Meadow et al., 1993; Matatyaho & Gogate, 2008), and timing of both (Gogate et al., 2000, 2006). For example, Maganti & Gogate (2013) examined the use of gestures in naming objects in mother-infant dyads and showed that mothers of both term infants and preterm infants adjusted their gestures when naming objects to their infants. Specifically, they used more gestures when naming objects to preterm infants than to term infants. This suggests that younger infants benefit from increased use of word-gesture combinations when their word knowledge is low compared to older infants. Indeed, multimodal presentations of words, actions, and objects are often synchronized in early development. Gogate and colleagues (2000) showed that by the middle of the first year of life, 76% of all interactions between infant and caregiver are aligned presentations of multimodal input, including words, objects, and actions (i.e., the object is clearly visible to the infant when it is named and moved), and these multimodal presentations remain synchronized 59% of the time by the middle of the second year of life.

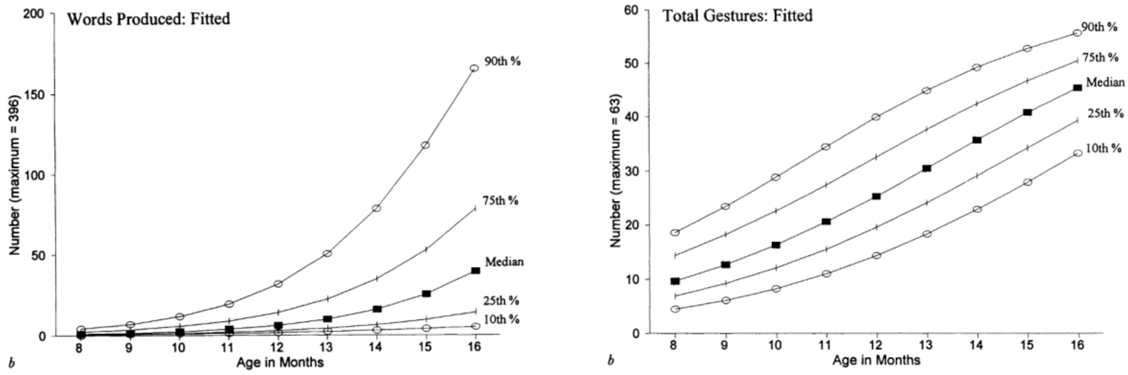
Actions facilitate word-object associative learning

Actions are therefore salient and dominant cues in input, often appearing alongside words, and are of great importance in cognitive development. Communication through the hands is an important component of learning, and infants encode and reproduce observed actions even before they speak their first word. For instance, once infants begin to develop a sense for the use of actions around 3 to 4 months of age (Zaadnoordijk et al., 2020), actions become an important source for infants to relate to others, so they relate even more to the actions they observe than to the faces of their interaction partners (Bahrick et al., 2002; Bahrick & Newell, 2008). Because gross motor movements are easy to perform and memorize, infants find it easy to understand and perform manual actions, thus placing less demand on working memory capacities and fine motor skills, such as tongue and mouth coordination in early word learning (Özçalışkan & Goldin-Meadow, 2005).

Infants use actions to communicate before they produce their first word, see Figure 7. For example, infants initially use deictic gestures, such as pointing to objects, and by 10 months of age they produce their first iconic gestures to communicate relational meanings, such as waving their hand in greeting (e.g., Iverson et al., 1994). Toward the end of the first year of life, combinations of actions and words in input become particularly important for infants' progress in expressive communication. For example, meaningful relationships between action and speech (e.g., "daddy" + point to cookie), rather than the simple presence of actions in naming events (e.g., "daddy" + point to daddy), predict the emergence of two-word utterances (e.g., Iverson & Goldin-Meadow, 2005). Thus, actions and gestures are precursors in early communication and complement language to express what infants intend but cannot yet put into words. Actions, then, can emphasize semantics in novel word-object mappings through their full-fledged gesture "vocabulary" and are undoubtedly a great help in the word learning process.

Figure 7

Early developmental trajectory of word and gesture production



Note. Number of words on the CDI form that infants growing up in an American-English speaking environment produces as a function of infant age in months. Data and graph taken from (Fenson et al., 1994).

The data from the above studies show that caregivers naturally adapt their gestures when naming objects to make the meaning of a word easily accessible to their infant. Actions and words therefore often occur together in the information infants receive. Actions are hence frequent and salient cues in input and facilitate word learning by complementing linguistic cues with hands. This makes it easier for young children to communicate with others and to encode the communicative intentions of others.

Actions

Early action-object associations

When objects are repeatedly moved in a particular way, or when objects exhibit an effect or function when moved, distinct action-object patterns emerge that the infant internalizes and recognizes in clearly visible conditions in the environment. Infants readily observe frequent action-object linkages and action effects elicited by others (e.g., flipping the light switch turns on the light or pulling the string of the music box triggers a song), leading them to discover objects from a dynamic visual perspective that allows them to communicate a goal or function unique to the action-object relationship. For example, at 6 months of age, infants predict action goals of individual objects: When the mother holds a cup, the infant is more likely to subsequently look at the mother's mouth than at any other part of the visual scene, even before the cup reaches the mouth, and the same is true for looking at hair when the mother holds a hairbrush (Hunnius & Bekkering, 2010). Similarly, 5- to 9-month-old infants in Woodward's (1998) study paid attention to the target a person was reaching for longer than to the actor itself. This suggests that infants are able to anticipate and predict the behavior of others even before they imitate action sequences themselves.

Actions performed on objects, then, are movements often presented in sequences (Monroy et al., 2017) that infants experience from different angles, with different lights, and sometimes even at different speeds. The perceptual dynamics that actions evoke are therefore highly variable, so unlike words, actions cannot serve as fixed referents for objects in the associative learning process. Although it is easier for infants to encode gross motor movements than specific phonological features of a word they cannot see or touch, actions may look a little different each time infants observe another person performing them. If the perceptual variability underlying action-object processing primarily facilitates infants' recognition of individual movement patterns, then associative learning of actions and objects should be based in large part on perceptual processes, so that infants integrate action associations from the bottom up (perception enables learning).

Actions as cues to visual category formation

Infants have been shown to categorize objects based on the type of movement that objects themselves elicit and the object functions performed on the objects that produce an audible or visible effect. Below, I summarize studies showing that infants as young as 10 months categorize objects based on movements and words, and 14-month-old infants categorize objects based on object functions and words only when the words are also accompanied by functions.

Object motions

A recent study by Sučević and colleagues (2021) examined the role of motions and labels in category learning and whether 10-month-old infants use these two types of information to categorize objects. Infants were first presented with a series of novel objects that differed in shape, color, and motion, while infants perceived the objects either silently or paired with a label or motion. In all 3 conditions, infants were able to learn the categories of the objects based on the information they received. However, only infants in the labeling condition showed the standard novelty preference indicative of category formation at test, whereas infants in the motion condition showed some evidence of category learning based on first-look analyses, and infants in the silent condition showed novelty preference only after several test trials. It could be that infants in the motion condition looked at the object from the category of objects that moved during training for longer or as long as the object from the unfamiliar category on test because the movements increased infants' attention. Detailed analyses of infants' gaze behavior during learning showed that both labeling and motion directed infants' gaze to the visual commonalities of objects with perceptual overlap, but silent presentation of objects did not. These results suggest that object motion highlights commonalities among objects with perceptual overlap and stimulates category formation similar to words, although the motion effect was not stable across analyses.

Object functions

Functional cues to object categories have been shown to be even more robust in early development. In Booth & Waxman's (2002) study, 14- and 18-month-olds were shown different objects with perceptual overlap either silently or accompanied by a functional cue (e.g., a hook attached to the object could be hung or swung) or word cue (“dax”). At test, infants were presented with two novel objects, one from the familiar category and one from an unfamiliar category and asked to select one that resembled the objects they had seen during familiarization and that was also the target object. 18-month-old infants showed category learning in all 3 conditions. 14-month-old infants showed robust category formation when the objects with perceptual overlap were presented in silence or with a functional cue, but not with a word cue. Thus, functional movements are likely to be important cues for category formation and may even be more robust than labels in object categorization at this age. Indeed, when infants were presented with both functional and word cues in the presence of similar-looking objects at 14 months, they generalized these objects and discriminated another member of the category from another object at test. Thus, infants categorized objects in silence, and on the basis of functional cues in the absence or presence of word cues. It is possible that functional cues are more salient to infants than words, perhaps because functions convey a goal for the movement performed on objects, whereas words are arbitrary in the way they initially refer to physical objects.

In contrast to most of the above studies of categorization, however, infants in Booth & Waxman (2002) were asked to select the object from the familiar category rather than interpreting longer looking times at the object from the unknown category at test as evidence for category formation. Thus, the functional effects might be related to the study design when the target object at test was the familiar object, because infants have preferences for novel or familiar cues in different modalities such as auditory and visual. Emberson and colleagues (2019) have shown that infants prefer familiar visual stimuli to novel stimuli and novel auditory stimuli to familiar stimuli. Therefore, it could be that the infants in Booth & Waxman's (2002) study preferred the object from the familiar category at test when the objects were presented with a functional movement but not a word, so that the rationale of the design gave the functional cues an edge over the word cues. Alternatively, it could be that 14-month-old infants who were presented with both functional and word cues with the object had a learning advantage because the simultaneous presentation of word and action cues provided infants with redundant information that can enhance their learning of the multimodal information.

Rationale of Study I: Temporal dynamics in early word-action-object learning

Intersensory redundancy

Indeed, when information is presented simultaneously in more than one sensory modality, such as vision and audition, it provides redundant information that can be used to improve infants' learning. For instance, Gogate & Bahrick (2001) investigated the role of intersensory redundancy in 7-month-old infants' memory for arbitrary syllable-object relations. They presented infants with pairs of objects and syllables, either in an audio-visual synchronous condition where the object and syllable were presented simultaneously or in an audio-visual asynchronous condition where the object and syllable were presented with a temporal offset of two seconds. Infants in the synchronous condition remembered the syllable-object pairings better than infants in the asynchronous condition, suggesting that the simultaneous presentation of auditory and visual information enhanced infants' learning for the arbitrary syllable-object relations. Thus, these results indicate that infant learning is enhanced when information is presented synchronously, i.e., redundantly across multiple sensory modalities. It is possible that redundant presentation of multimodal word and action cues makes it easier for infants to recognize non-redundant information in the input (Bahrick et al., 2004; Bahrick & Lickliter, 2000).

Building on the hypothesis of intersensory redundancy, Gogate and colleagues (2001) present a theoretical framework for understanding how infants acquire word comprehension skills through the integration of sensory experiences. They propose that word comprehension results from the dynamic interaction between the sensory input (visual, auditory, tactile, etc.) that infants receive and their developing motor and cognitive abilities. Thus, from the perspective of this dynamic system, infants learn to associate words with objects through a process of cross-modal integration as they combine information from different sensory modalities into a unified representation.

When actions accompany word-object associations in experimental tasks, infants do not show learning

However, the robust learning patterns described in the word learning literature are distorted when infants have the opportunity to link information from more than one modality (e.g., words and actions) in a laboratory task. Although experiments using cues from multiple modalities more closely match infants' natural learning environment, infants show impoverished learning of novel word-object pairs in such tasks. For example, Eiteljoerge and colleagues (2019) directly compared the learning of word-object and action-object pairs in 1-, 2-, and 3-year-old infants and adults. Participants were first presented with two novel word-object and action-object mappings and then tested on their recognition of these mappings. At test, they were shown the two objects from the familiarization phase side by side, which then disappeared to present a word or an action prime in the absence of the object (i.e., participants heard the name for one of the objects or saw a movement performed on the objects during the familiarization phase). Following the prime, they were again presented with the two objects from the familiarization phase side by side. Increased attention to an object previously associated with either the prime label or the action would indicate recognition and learning of the word-object or action-object association. They found that 1-year-old children learned action-object associations but not word-object associations; 2-year-olds learned neither association; and 3-year-olds learned word-object associations but not action-object associations; whereas college-age adults learned both. Other studies found similar results: When 15-month-olds saw word-object associations accompanied by actions, they showed no recognition of word references at test (Puccini & Liszkowski, 2012). Such findings suggest that young children attend to input from a single preferred modality when presented with information from two modalities (e.g., Lewkowicz, 1994; Robinson & Sloutsky, 2004; Sloutsky & Napolitano, 2003). It is possible that the above studies, particularly that of Eiteljoerge and colleagues (2019), did not find robust learning of words in such multisensory environments because the words and actions for objects in the experimental task were not synchronized, which may have disrupted children's learning of the associations.

Intersensory redundancy and temporal synchronicity of word-action-object triads

In Chapter 2 below, I show that manipulating the temporal patterns underlying word-action-object triads does indeed alter infants' word-object and action-object learning patterns in early development, with synchronous but not sequential presentations reinforcing learning of word-object and action-object associations by 2 years of age. Our data provide evidence that the impoverished learning of word-action-object triads reported in the literature was reversed when 2-year-olds were presented with these cues synchronously, but not when these cues were presented sequentially or to younger children (1-year-olds). It is possible, therefore, that synchronous presentation of words, actions, and objects enabled 2-year-olds to recognize relevant, non-redundant information in the input, whereas they were unable to do so when these cues were presented sequentially. Thus, intersensory redundancy not only plays a role in the processing of multisensory information, but also enables the learning of these triads in experimental settings where learning of multisensory cues is not otherwise observed in early development.

Rationale of Study II: Arbitrariness of cues in category formation

Data on word and function effects in object category formation tasks showed that 14-month-olds categorized objects with perceptual overlap when those objects were presented with a functional cue, both in the presence and absence of word cues (see the section on actions above for a summary of Booth & Waxman's (2002) study). Word cues, however, were informative of category membership only when objects were additionally accompanied by functional cues. Such differences could be due to the fact that mediated functions are often meaningful in the way they refer to objects, whereas words are arbitrary in the way they initially refer to physical objects. In the following chapters, I compared word cues to action cues rather than to functions, since actions can be reconciled with the arbitrariness of word-object associations. That is, the sound of a word and the movement underlying an action do not initially explain which object these cues refer to until these cues have been repeatedly and overtly assigned to their referents, or at least in this context we will treat words and actions as such. Here, arbitrary actions are dynamic sequences performed on an object that have no referential meaning in the movement they perform, such as turning the object upside down and back up again, which produces no effect or intended goal like functional cues.

Incidentally, when I speak of associative and categorical learning of words in the following studies, I refer to novel nouns as labels for objects and object categories. Infants learn nouns before they learn other word types such as verbs and adjectives, and nouns are reliable referents in novel word-object mappings (e.g., Booth & Waxman, 2009). We used pseudonyms as fictitious object names (e.g., "Tanu" or "Loeki") to compare the learning of these names with previous studies examining the learning of words, actions, and objects (Eiteljoerge et al., 2019) and to ensure that the words infants heard in the presence of new objects were as novel as the arbitrary actions introduced during the familiarization phases. Comparing pseudonyms to arbitrary actions that match in arbitrariness therefore allows us to compare modality-specific input effects, with a specific target for words and actions.

This is what I did

Study I: Chapter 2

I examined infants' word and action learning patterns in multisensory environments (where objects, actions, and words are present) because previous studies have shown that infants and young children show reduced word learning for objects when words were accompanied by action cues in a laboratory task (Eiteljoerge et al., 2019; Puccini & Liszkowski, 2012). To rule out the possibility that infants showed no prior learning because the presentation of stimuli in such laboratory tasks was not synchronized, as is the case in infants' natural learning environment, I revisited associative learning of words, actions, and objects from the perspective of temporal synchronicity. Specifically, the study aimed to gain a broader understanding of the mechanisms underlying word-object and action-object learning by manipulating the timing of these combined cues. That is, I systematically investigated the timing of the co-occurrence of word and action cues along with a single object using eye-tracking in a preferential looking task and pupillometry in a violation of expectation paradigm in 1- and 2-year-olds.

As suggested by the looking time data, I found evidence of associative learning of words with objects and of actions with objects in 2-year-olds when the stimuli were initially presented synchronously but not when they were presented sequentially. In contrast, the 1-year-olds did not show learning success in either temporal condition. This could be because repetition during familiarization was insufficient for learning or because younger infants are more dependent on additional social factors involved in multisensory processing. Pupil measurements indicated no learning in either age group, perhaps because violations were more difficult to detect compared with previously seen word-action-object mappings. It is also possible that the pupil measures did not support looking time results because the violation of expectation paradigm was always presented in the second part of the experiment to avoid that mismatch demonstrations of these associations override previously learned associations, so that learning of the cues during familiarization was transient until this later part. The results thus indicate that infants either learned to associate both words and actions with objects or neither information, suggesting that word-object and action-object do not differ in early development or are modulated differently by the temporal features with which these cues were presented. Thus, the study shows that word and action cues in associative learning share the same sensitivity to the temporal dynamics underlying the presentation of such cues, influence learning in similar ways, and may share common structures underlying learning.

Study II: Chapter 3

In Chapter 3, I present data on word and action effects in visual category formation at 1 year of age using basic level objects (all stuffed animals with perceptual overlap within the familiarized category). Specifically, I used the novelty preference paradigm to examine the extent to which action and word cues similarly influence object categorization. If infants use actions similarly to words to generalize objects and form an exclusive category, then words should be no more special than arbitrary actions in categorization. During the familiarization phase, there were 3 groups, i.e., 1-year-olds see 8 objects with considerable perceptual overlap, one at a time, paired either with a single cue (either an action or a word) or with both cues (word and action) in synchrony. At test, all infants see the same two test objects in silence. They show a novel exemplar of the category they just learned next to a novel object from an unfamiliar category, the latter being the target object because this object is relatively novel compared to the exemplar within the category. I examined the effect of word and action cues in basic object categorization using two statistical approaches.

First, we used a Sequential Bayesian Factor (SBF) analysis to examine the probabilities that H1 (novelty preference for the target object at test) and H_0 (equal looking at both objects at test) were true given the data. We set an upper threshold of 3 for moderate evidence for H_1 and a lower threshold of 1/3 for moderate evidence for H_0. I found moderate evidence for H_0 in the action and word-action conditions and anecdotal evidence for H_1 in the word condition with 30 infants in each group. Second, I used a generalized linear mixed model (GLMM) and found evidence for categorization indexed by novelty preference at test in the word but not the action or word-action condition. Thus, the results suggest an advantage of words in early object categorization, although this was not stable across analyses.

So, there is some evidence that words cue basic object categorization, but arbitrary actions do not. Moreover, when words are accompanied by arbitrary actions, infants did not form an exclusive category for the training objects. On the one hand, the actions may have been too complicated, such that they distorted the encoding of the objects and caused higher processing costs compared to words, which affected category formation. Thus, the cognitive load might have been higher when infants saw the objects move with and without words. On the other hand, words might have facilitated object categorization because word processing involves additional levels of cognitive control that play a role in category learning. If cognitive control and language processing are intrinsically linked, then naming similar-looking objects can promote such learning, which other arbitrary cues cannot. However, it is not possible to conclude that actions interfered with category formation or that words promoted it, because we did not compare these eye-tracking data with silent and still presentation of objects. If infants in such an additional condition in which no cue was given showed signs of category formation at test, then infants in the word condition might simply have ignored the word while looking at the objects, so that words neither facilitated nor hindered category formation.

Study III: Chapter 4

The purpose of this study is to examine the relationship between category learning and word learning in early development. Previous studies suggest that these two processes are interdependent. For example, words have been shown to facilitate the formation of categories (see Introduction and Chapter 3), and recently a study has shown that knowledge of categories promotes the learning of category names (J). To further investigate the relationship between these processes, two groups of 40 12-month-old infants participate in the study, completing a category learning task and a word learning task in different orders. The first task is designed to test whether the infants learn word-object associations or form object categories, and the second task is designed to test the influence of learning from the first task (i.e., word-object knowledge tested for category formation or category knowledge tested for word-object learning). If the two processes are independent, the order of the tasks should not affect infants' learning. However, when they are linked, infants' success on each task depends on their learning on the first task. The results of this study will provide further insight into the relationship between word learning and category formation in early development.

Due to contract restrictions related to the COVID -19 pandemic, I initially conducted the study online. However, because there were differences between the extent to which infants attended to the stimuli online and in the laboratory, as indicated by the data presented in Chapter 3, I moved the study back to the laboratory once regulations again permitted in-person testing. Data collection is still ongoing, so I will not be able to present the results of the study until the dissertation is submitted. However, I will briefly discuss the significance of the study in the discussion and therefore include this study as a brief report in this thesis.

Methods

Pandemic-related changes to the original plans for the thesis

Note that due to contact restrictions related to the Covid-19 pandemic, eye-tracking data collection in the laboratory was suspended for approximately 16 months (March 2020 - May 2021). During this time, I switched data collection in the category study to an online format and manually coded video recordings of infants looking behaviour instead. In this online format, infants sat on their caregiver's lap and participated from home using a laptop with a camera and speaker. For the study presented in Chapter 3, we collected online data from more than 160 infants before moving the study back to the laboratory to replicate the online results using the eye-tracker. The analyses show that there are considerable differences in the extent to which infants attend to the objects on the screen presented online and in the laboratory, see Figure 20. Therefore, the data presented in Chapter 3 include only the eye-tracking data collected in the laboratory. Although the online dataset included a control condition in which only objects but no additional cues were shown, I was unable to complete data collection for this condition in the laboratory by the time the dissertation was submitted. In the near future, the data collected in this project will be used to answer two different research questions, each of which will be formulated in a separate article. One article will highlight the differences in online and laboratory data collection with respect to infants' looking behavior in visual category formation tasks using data from the word condition, as this is the only condition for which I have found evidence of an effect on category formation. The second article addresses the initial research question, namely, the extent to which word and action cues individually and in combination modulate category formation at 1 year of age.

Participants

I recruited 1- and 2-year-olds for the associative learning study in Chapter 2 to investigate whether temporal manipulation of word-action-object triads allows infants to learn these cues to objects, which has not been the case in previous studies examining the learning in conjunction with action cues. Therefore, to allow developmental comparisons between these studies, we tested infants of the same age group as in (Eiteljoerge et al., 2019).

For the categorization study in Chapter 3, we tested 1- and 2-year-old children online to compare word and action effects across tasks and development. We hence completed and coded data sets for both age groups in who participated in the online format. When we brought the study back to the lab to replicate the online results, we tested only the 1-year-olds to allow for study completion by the end of the PhD and to spare resources. The decision to examine the effects of word versus action cues on basic categorization skills in the younger age group was motivated by previous research indicating the robustness of word-based effects in this population, even at 1 year of age. Given this robustness, it was deemed appropriate to replicate these findings and compare them to the effects of action cues in this age group. In case we did not find effects in any of the conditions with 1-year-olds, we planned to increase the age window and test 2-year-olds to get a more complete picture of word and action effects in early development.

Materials

Objects

I used novel objects as references for novel words and actions in both the associative learning study (Chapter 2) and the category learning study (Chapter 3). The objects were either visually changed from their original appearance or handmade to ensure that the objects were new to the infants. The two novel objects in Chapter 2 were kitchen utensils, namely a fruit press and a dough roller. The fruit press was stuffed with white cotton so that it was not transparent, and the tips of the dough roller were dyed blue, see Figure 9. During the preferential-looking paradigm, infants saw another transparent object in the form of a Plexiglas tube during the action primes. The latter object served to demonstrate the action that had previously been associated with one of the novel objects during familiarization. I could not use one of the objects to demonstrate the action during the priming phase because the object would have revealed information about the object-action association prior to the recognition phase, in which infants' learning about these associations was tested.

In Chapter 3, 10 novel objects with perceptual overlap were used to test infant's category formation for these objects. All infants saw 9 stuffed toys from the same perceptual category (8 objects during training, 1 object at test) and 1 stuffed toy that did not belong to this perceptual category (see Figure 14). The objects were selected from a larger sample based on similarity ratings by college-age adults (see Chapter 3 for details on the similarity study).

Words

I used noun pseudonyms as auditory stimuli for novel objects to avoid the possibility that infants already knew a name. A 28-year-old woman produced the auditory stimuli in infant-directed speech as a native German speaker. In both studies, the auditory stimuli for the training phase consisted of blocked passages introducing the novel names, which were either "Tanu" or "Loeki". Thus, for the groups that saw an object while hearing a name for that object in the studies, the label was presented with a linguistic interjection, i.e., "Ah! Ein Tanu! Oh, ein Tanu! Wow, ein Tanu!" (Ah! A Tanu! Oh, a Tanu! Wow, a Tanu!).

Actions

I used action sequences that were arbitrary in the way they acted upon the object. That is, the action conveyed neither a goal nor a visible effect. The actions performed on the objects in the 2 experiments reported here were identical, that is, the objects were either turned upside down, which provided visual information about the bottom of the object, or rotated around their own axis, which provided visual information about the back of the moving object, see Figure 15.

Procedure

Families visited us at the Wortschatzinsel Baby Lab at Gossler Strasse 14 in 37073 Goettingen to participate in a study. Upon arrival, the tester greeted the family and played with the infant while caregivers completed a consent form and agreed to inform the lab team of any corona infections diagnosed shortly after their visit. When the infant became comfortable with the tester or lost interest in the game, the tester moved on to the eye-tracking study.

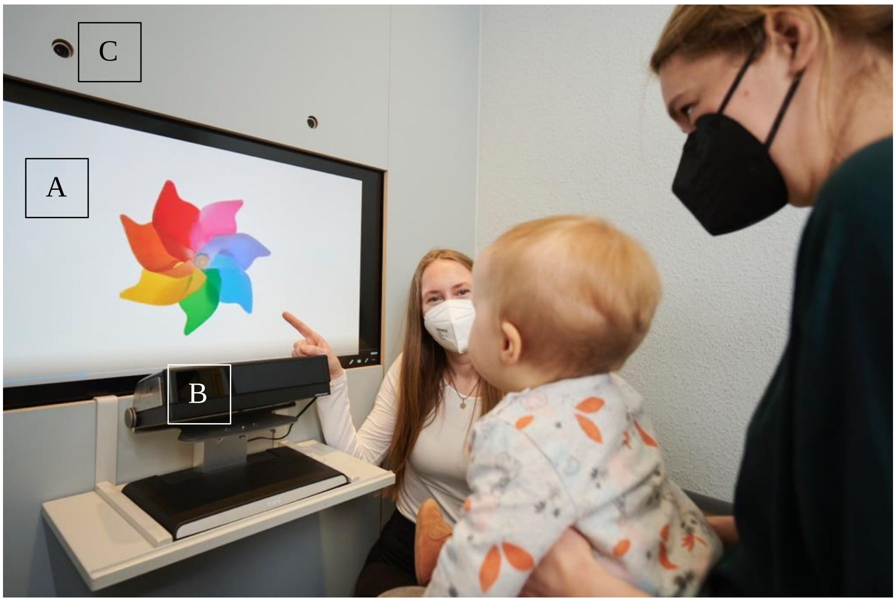
Eye-tracking

Eye-tracking is a research technique used to study visual attention and perception. In early developmental research, it involves measuring eye movements and fixations, as well as pupil changes, while infants view various visual stimuli, such as pictures or videos. The eye tracker consists of a camera positioned near the infant's head that tracks the reflection of light from their eyes as they look at the stimuli displayed on a screen. The eye tracker records the position, duration and frequency of fixations (periods of sustained gaze) and saccades (rapid eye movements) in response to the visual stimuli (e.g.,Gredebäck et al., 2009).

The child was seated either in an infant carrier or on the lap of his caregiver at a distance of 55 to 65 cm in front of a 92 x 50 cm TV screen (A), see Figure 8. The Tobii X 120 remote eye tracker (B), placed underneath the TV screen, recorded gaze data at 60 Hz, and Tobii Pro Labs software was used for stimulus presentation. Two speakers mounted above the TV screen presented auditory stimuli, while visual stimuli were presented on the TV screen. Two cameras (C) located directly above the TV screen provided online recordings of the child and were used to track the child during the experiment. We used a five-point grid calibration (one point in each corner and one in the center of the screen) in Tobii Studio

Figure 8

Eye-tracking booth in the Baby lab in Göttingen



Note. The tester (me) points to the stimulus on the screen to draw the infant's attention to the screen and facilitate calibration. Before the experiment, the tester leaves the booth so as not to interfere with the infant's attention during eye-tracking. Marina Proft (infant's mother on the right side of the picture) gave permission to use the image to demonstrate the eye-tracking procedure. The image was published in connection with the interview "Mehr Wörter als erwartet" (*more words than expected*) in Baby & Familie in July 2022. In the interview, I talked about recent findings on the development of children's vocabulary during the pandemic.

Paradigms

Preferential looking

The preferential looking paradigm is a commonly used method for studying infants' associative learning (e.g., Aslin, 2007), such as word-object and action-object recognition in learning. In the study reported in Chapter 2, two objects are presented, each paired with a specific word and object. After this initial learning phase, the infants are presented with the two objects simultaneously, i.e., silently side by side on the screen, while the eye tracker measures the time the infants spend looking at the two objects.

By comparing the time infants spent looking at the two objects, we were able to infer infants' preferences for the objects without any additional cues. After three seconds, the pictures of objects disappeared, and infants instead saw an action prime or heard a word prime. That is, infants either saw one of the actions from the familiarization phase performed on an unfamiliar transparent object or heard a word from the familiarization phase while looking at a black screen. The two objects, presented silently side by side, were again shown on the screen with the expectation that this time infants would look longer at the object with which the previous action-prime or word-prime was associated. Thus, we compared gaze times to both objects before and after the prime phase to examine the learning of word-object and action-object associations.

Violation of expectation

The violation of expectation paradigm used in the associative learning study described in Chapter 2 is a widely used method in infant research that we have used to study infants' learning of word-object and action-object associations. The basic idea behind this paradigm is that infants' pupils dilate during events that violate their expectations as index of surprise and hence learning of the associations they have previously seen compared to events that are consistent with their expectations (e.g., Zhang & Emberson, 2020). Specifically, following word-action-object familiarization, we successively presented infants with two events. One matched the word-object and action-object associations presented during habituation and thus matched their expectations regarding the occurrence of these associations. The other one showed the object with action and word cues that differed from those presented during familiarization and thus conflicted with their expectations (Hepach & Westermann, 2016). If infant's pupils are larger during the latter event than during the familiar event, this is taken as evidence that children have recognized a violation of their expectations and is actively trying to make sense of the situation (Jackson & Sirois, 2009). Thus, pupil dilation has the potential to act as a measure of learning because changes in pupil size are associated with unexpected psychological events.

Novelty preference

As in most of the categorization studies reported above, we examined infants' category learning using looking time measures in a novelty preference paradigm. In the study reported in Chapter 3, 1-year-olds initially saw 8 pictures of stuffed toys that overlap in perception (i.e., similar shape and facial features within a category), one at a time. During this habituation phase, infants' engagement is usually high in the beginning when the objects are novel, and their attention to the objects usually decreases after they have seen some members of the category (Althaus & Westermann, 2016; Sirois & Mareschal, 2002). This was the case for infants in the word and action condition, but not in the word-action condition, so infants in the latter condition did not show habituation to the familiarized category. At test, infants see two objects side by side, an unseen member of the familiar category and a novel object outside that category. When infants successfully habituate to the category (as infants did in the word and action condition), they are expected to look longer looking times for the out-of-category object simply because it is more novel and thus more interesting than the in-category object. Longer looking times for the novel object from the unfamiliar category are usually taken as an indicator of successful generalization of members from the familiar category and category formation that excludes the object from the unfamiliar category.

Analyses

Generalized Linear Mixed Models (GLMM)

GLMMs are statistical models that extend the generalized linear model (GLM) to account for correlation structure in data collected from experimental units, i.e., random effects structure. In a GLMM, the response variable may be continuous, binary, or count data. The models reported here are Gaussian and beta models with proportional looking values bound between 0 and 1, that is, looking times cannot exceed the stimulus display on the screen. The model contains fixed effects, i.e., factors that are held constant across all sources or units. Random effects are factors that vary across sources or units. The model also includes an error distribution and a link function that relates the response variable to the predictor variables. The mixed part of the model refers to the inclusion of both fixed and random effects, which allows the modeling of correlation structures within a subject or group.

Full-null model comparison

To avoid cryptic multiple hypothesis testing, the fit of a full model with all variables is compared to a null model that excludes the predictor of interest. Cryptic multiple hypothesis testing is a statistical problem that occurs when multiple hypotheses are tested simultaneously without adjusting the significance level or accounting for the fact that multiple tests are being conducted (Forstmeier & Schielzeth, 2011). This can lead to an increased risk of obtaining false positives, where significant results are discovered by chance rather than reflecting a true relationship between the variables under study. Cryptic multiple hypothesis testing can therefore lead to overestimated effect sizes, where the size of the observed effect is inflated due to random variation, making the results of the analysis unreliable. Full-null model comparisons therefore allow the unique contribution of each variable to the model to be assessed while controlling for the other variables (Gelman & Hill, 2007). This approach can reduce the risk of random overestimation of effect sizes and is commonly used in regression analyses and other statistical models.

The "full model" includes all possible predictors and the interactions between predictors and aims to explain the relationship between the response variable and all available predictors in the data. It is often used as a benchmark to compare with simpler models that include fewer predictors. By including all predictors and interactions, the full model provides a comprehensive understanding of the relationship between the response variable and the predictors. A null model is tested against the full model to assess whether the predictors are statistically significant and whether they improve the overall fit of the model. The null model is a simplified statistical model derived from the full model that assumes the absence of certain predictor variables that are critical to the hypothesis being tested and represents the null hypothesis that there is no effect. If a full model does not provide a significant improvement in fit compared to the null model, this may indicate that the additional complexity of the model is not necessary to explain the variation in outcome variables.

We compared the full models with their respective null models using a chi-square test to determine whether the predictor of interest contributed significantly to model fit and whether it should be retained in the final model. The chi-square test compares the observed data to the expected data under the null hypothesis that there is no difference between the compared groups. The test statistic is calculated as the sum of the squared differences between the observed and expected values divided by the expected values. If the test statistic is greater than a critical value determined by the chi-square distribution with degrees of freedom equal to the number of categories minus one, then the alternative hypothesis is likely to be true at some level of significance (here alpha equalled 0.05) and it can be concluded that there is a significant difference between the compared models. The chi-square test is used with the function ANOVA, a statistical function commonly used in statistical software packages such as R (R Core Team, 2022) to perform an analysis of variance. When comparing the full model to the null model, the function ANOVA compares the models by testing whether the difference in variance is statistically significant, i.e., calculating the difference in variance between the models.

Regression coefficients

The regression coefficients of the full model are used to estimate the strength and direction of the relationship between the independent variables and the dependent variable in the regression model and to make predictions about the dependent variable based on the values of the independent variables. Each coefficient represents the expected change in the dependent variable associated with a one-unit increase in the corresponding independent variable, while holding all other independent variables constant. In other words, the coefficients explain the average effect of each independent variable on the dependent variable, considering the effects of all other independent variables in the model. A positive coefficient means that the dependent variable is expected to increase when the independent variable increases. In contrast, a negative coefficient means that the dependent variable is expected to decrease if the independent variable increases.

Overall, the regression coefficients are used to analyse and predict the relationships between the independent variables and the dependent variable in the full model. The GLMMs used maximum likelihood estimation (MLE) to estimate the regression coefficients for the critical fixed effects along with other model parameters. To do this, we used the lmer function in the lme4 package in Chapter 2 and the Template Model Builder (TMB) in Chapter 3. TMB is a software package for fitting statistical models that enables the use of complex models with many parameters by automating the estimation process and providing efficient optimization of the likelihood function. MLE is a statistical method used to estimate the parameters of a model by finding the parameter values that maximize the likelihood of the observed data. In the context of regression models, MLE is about finding the parameter estimates that make the observed data most likely given the assumptions of the model.

Sequential Bayesian Testing

Given the lack of a learning effect in the online dataset, we conducted sequential Bayesian testing in Chapter 3 to determine the optimal sample size. This was done to assess the plausibility of an inadequate sample size as a cause of the failure to observe the effect. Sequential Bayesian testing is a statistical method for making decisions based on data collected over time. It involves updating the probability of a hypothesis after each observation of new data and then planning on how to proceed with data collection based on the updated probability. The basic idea of sequential Bayesian testing is to start with a prior probability distribution for a hypothesis and then update that distribution as new data become available (Lecoutre & Poitevineau, 2011). This can be done using Bayes' theorem, which states that the probability of a hypothesis given the data is proportional to the product of the prior probability and the probability of the data given the hypothesis (Broemeling & Good, 2007; Gelman et al., 2014; Stefan et al., 2019). As new data are observed, the posterior probability distribution is updated, and the decision can be made based on the updated probability.

Thus, sequential Bayesian testing allows to either to accept or reject a hypothesis, or to continue collecting data after every infant tested. We calculated a Sequential Bayes Factor (SBF; Mani et al., 2021; Schönbrodt et al., 2017) after collecting data from individual infants to examine the likelihood of the data under the alternative hypothesis H_1 (infants look longer at the object from the novel category than at the one from the familiar category) compared with the null hypothesis H_0 (there is no difference in looking at the object from the familiar category than at the one from the new category). The BF was given minimum level of evidence (1/3 < BF < 3; i.e., moderate evidence for either hypothesis), such that exceeding these BF thresholds in any of the experimental groups (either action, word, or word-action) would provide evidence for the alternative or null hypothesis.

Outlook

In the following studies, I examine the effects of words and arbitrary actions in a multisensory associative learning task and in a category formation task. In the first task, infants are presented with words, actions, and objects either synchronously or sequentially. Subsequent learning of word-object and action-object associations is tested using preferential looking and pupil changes in a violation of expectation paradigm. Because infants are used to the synchronous presentations of co-occurring actions during word-object presentations in their social learning environment, we expected that infants associate words and actions with objects in the synchronous condition, but not in the sequential condition. The latter condition served to replicate the results of previous studies that had shown that infants do not learn words when accompanied by actions during learning tasks were shown asynchronously in the laboratory. These predictions were fulfilled at 2 years of age but not at 1 year of age, suggesting that processing words, actions, and objects at this stage of development comes at an additional cost.

In the second study, 1-year-olds were first presented with several objects that share many perceptual features and were accompanied by either a word, an action, or both. Then, their visual novelty preference is tested for an object outside the category that is presented next to another member of the familiar category. There is some evidence that infants formed a category at test when the objects were first introduced with a word, but not when objects were introduced with an action or both cues. While word-action cues may have overloaded infants' cognitive capacities similar to the infants in the study above, action cues may have been too complicated for the infants show evidence for category formation. The category formation task is arguably more complex in that infants recognize similarities between objects and generalize these features to unify them into a common mental representation. Here, words are, therefore, likely to overcome cognitive limitations that actions cannot, giving words an edge in concept formation. I also discuss the rationale and design of a third study that explores the connection between word learning and category formation. While the results cannot be presented until the submission of this thesis, this study has the potential to enhance our understanding of the relationship between these two aspects in early development.

Chapter 2

Better in sync: temporal dynamics explain multisensory word-action-object learning in early development

Ricarda Bothe[1,3], Sarah Eiteljoerge[1,3], Léonie Trouillet[2], Birgit Elsner[2] & Nivedita Mani[1,3]
[1]Georg-August University Göttingen, [2]University of Potsdam, [3]Leibniz Science Campus

Author statement

The following article has been accepted as a Stage I registered report at INFANCY on January 22nd, 2022. The Stage II manuscript was resubmitted on January 23rd, 2023 and is currently under review.

The authors confirm contribution to the paper as follows:

Study conception: Ricarda Bothe, Sarah Eiteljoerge, Nivedita Mani

Study design: Ricarda Bothe, Sarah Eiteljoerge, Léonie Trouillet, Birgit Elsner, Nivedita Mani

Data collection and analysis: Ricarda Bothe

Interpretation of results: Ricarda Bothe, Birgit Elsner, Nivedita Mani

Original draft manuscript preparation: Ricarda Bothe

Draft manuscript revision: Sarah Eiteljoerge, Léonie Trouillet, Birgit Elsner, Nivedita Mani

All authors reviewed the results and approved the latest version of the Stage II manuscript.

Abstract

By the middle of the first year of life, infants recognize a handful of words and actions for physical objects. Studies examining the factors underlying such word-action-object learning have employed laboratory tasks where 1- and 2-year-olds are presented with either words or actions. The input that infants are naturally exposed to, however, is multisensory, with simultaneous presentation of objects, actions and words. Here, we examined how word and action learning occurs when infants are presented with information from linguistic-auditory and action-visual modalities simultaneously, and the factors that influence early learning in multisensory environments. Specifically, we examine the influence of temporal systematicity of word and action input on learning, while controlling for potential individual differences in early language abilities. Infants were trained on word-object and action-object associations as they were either synchronously or sequentially presented with novel objects, each associated with a novel label (e.g., "Oh, a Tanu!") and an arbitrary action (e.g., hand rotates the object). Following training, we tested 1- and 2-year-old's recognition of the trained novel word-object and action-object associations, using both preferential looking time and pupillary responses as indices of recognition. At 2 years, we found evidence for word-object and action-object learning when word-action-object training was synchronous but not when training was sequential, suggesting that word- and action-cues operate similarly with regard to the multisensory temporal dynamics in which they occur. At 1 year, we found no evidence for learning of the novel word-object and action-object pairings or modulation of learning by temporal systematicity. Together, these results show that temporal dynamics influence word-object and action-object learning similarly during the first 2 years of life, suggesting that shared processing mechanisms are equally sensitive to temporal structures in multisensory presentations.

1. Introduction

When infants hear a label in the presence of an object or see an action being performed on an object, they learn to associate this information and eventually form distinct word-object and action-object mappings. As infants draw such mental connections, they learn to structure complex and dynamic input from different modalities (i.e., movements and language), which characterizes an important milestone in infant's development of cognitive and communication skills (Swingley, 2009; A. L. Woodward & Markman, 1998). There is considerable information available regarding infant's acquisition of such understanding when presented with either word-object (Bergelson & Swingley, 2012b; Fenson et al., 1994; Taxitari et al., 2020; Tincoff & Jusczyk, 1999) *or* action-object associations (e.g., Falck-Ytter et al., 2006; Kanakogi & Itakura, 2011; Mirijamdotter & Somerville, 2005), i.e., in settings where they are either told the name of an object or shown an action that one can perform on the object. However, infants are often presented with information from multiple modalities simultaneously (i.e., actions and words for objects, henceforth multisensory settings, see also Broadbent et al., 2018) and might be able to simultaneously learn to associate both actions and words with objects. Such multisensory settings, which arguably approximate infant's natural language environment better (Gogate et al., 2006), reveal different patterns of learning of word- and action-object associations relative to multimodal settings in which only action-object or word-object pairings occur. In particular, while some studies suggest that multisensory input (e.g., inter-sensory redundant information) facilitates learning and retention of such associations at 7 months of age (Gogate et al., 2001; Gogate & Bahrick, 1998), more recent data suggests that infants show impoverished learning of word-object pairings in multisensory settings between 12 and 30 months (Childers & Tomasello, 2003; Dysart et al., 2016; Eiteljoerge et al., 2019; Puccini & Liszkowski, 2012; Yoon et al., 2002) .

Differences in the findings may either be explained by differences in the timing of presentation of the stimuli or reduced sensitivity in the paradigm used to detect learning. Against this background, across different paradigms – the preferential looking paradigm (e.g., Golinkoff et al., 2013) and the violation of expectation paradigm (VOE; Baillargeon et al., 1985) – we examine 1- and 2-year-olds' learning of word-object and action-object associations in multisensory settings at one- and 2 years of age with particular emphasis on the extent of shared variance in action-object learning and word-object learning across early development. Furthermore, in keeping with suggestions that temporal synchronicity has been shown to boost learning in multisensory settings in the early stages of development (Bahrick & Lickliter, 2004; Gogate, 2010), we examine how the temporal alignment of stimuli from different modalities may influence learning, while controlling for individual differences in 1- and 2-year-old's success at word- and action-learning. In the following, we will review the literature on word and action learning and discuss the role of temporal systematicity in early language and action acquisition.

Word-object and action-object learning

Research attempting to capture early word learning progress has generated a wealth of data on infant's ability to attune to spoken language input over the course of the first years of life (Fenson et al., 1994). This work shows that even infants as young as 6 months of age show recognition of highly familiar word-object associations such as *Mama*, *hand*, *foot* etc. (Bergelson & Swingley, 2012b; Tincoff & Jusczyk, 1999). By the time infants reach their first birthdays, they understand on average 50 words while their general receptive vocabulary further grows to over 150 words by 16 months (Fenson et al., 1994). This pattern of vocabulary development highlights young children's ability to successfully encode novel objects using linguistic cues from early on. This ability develops further with the average 2-year-old learning around two new words per day (Bloom, 1993) and producing new words immediately after hearing them. Such "fast mapping" of novel words to novel objects (e.g., Borgström et al., 2015; Swingley et al., 1999; Swingley & Aslin, 2000) emphasizes robust learning of newly presented word-object associations by 2 years of age. Nevertheless, in less naturalistic laboratory settings, while infants as early as 10 months have been shown to learn novel word-object associations, infants show limited effects of learning, retention and generalization of these newly-learned word-object associations at this young age (e.g., Fenson et al., 1994; Frank et al., 2020; Misyak & Christiansen, 2012; Pruden et al., 2006; Taxitari et al., 2020; Twomey et al., 2018; Woodward et al., 1994).

Action learning and action processing advances at a similar pace in infant development, for example, by 6 months of age, infants learn to map actions onto novel objects, recognize familiar action-object associations (Bahrick et al., 2002). Also by this age, infants anticipate the goal of a human grasp

(Cannon & Woodward, 2012; Falck-Ytter et al., 2006; Kanakogi & Itakura, 2011) and the goals of particular actions on familiar objects (i.e., cup is moved to mouth; Hunnius & Bekkering, 2010). These abilities develop considerably in the second year of life, with toddlers being able to learn the effects of particular trained actions as they relate it to their own actions by 15 months (Elsner & Aschersleben, 2004), and correctly interpret the intended goal of failed actions (Brandone et al., 2014; Brandone & Wellman, 2009; Meltzoff, 1995).

Furthermore, there is considerable evidence available showing that actions are extremely robust and salient features of the environment in the early stages of development: For instance, 5-month-olds discriminate and retain actions but not static images of human faces (Bahrick & Newell, 2008) and during the second year of life, young children produce significantly more words when combined with a movement or gesture relative to words that are not accompanied by such a salient action cue (Özçalışkan & Goldin-Meadow, 2005). Robust action recognition and retention and the relatively fragile word learning patterns during the first year of life suggest that actions may serve as a window into language learning, forerunning lexical and linguistic abilities (Capirci et al., 2005; Caselli et al., 2012; Özçalışkan & Goldin-Meadow, 2005).

Word-object and action-object learning in multisensory environments

The studies mentioned above have typically tested infant's learning of word-object and action-object mappings in isolation from each other. In their natural learning environment, however, infants are often confronted with complex linguistic and dynamic visual information as they encode relevant cues from the environment (Werker et al., 1998), setting the stage for successful communication (Gogate et al., 2000; Gogate & Hollich, 2010). In the following, we will address some of the literature focusing on the role of early word-object and action-object learning when rich information from both the linguistic and the action modalities are presented dynamically and simultaneously.

Early in development, there appears to be some benefit of being presented with information in such multisensory environments. For instance, 6- to 8-month-olds learn word-object mappings when actions and words were presented in a temporally synchronous manner and when infants attended to the presentation of the multimodal input (Gogate et al., 2006; see also Matatyaho-Bullaro et al., 2014;

Werker et al., 1998, albeit at 14 months). More recent work by Goldstein et al. (2020) shows that when 13-month-olds experienced objects with words in infant directed speech (IDS) and actions in infant directed motion (IDM; i.e., exaggerated movements such as shaking), infants learned words more reliably when the object was moving than when the object was stationary (see also Werker et al., 1998). Interestingly, this effect was not found at 15 months, suggesting that mechanisms underlying multimodal word and action learning crucially develop over the second year of life.

A recent study investigating the developmental time course of word and action learning in a multisensory context and directly comparing word-object and action-object learning in one-, two- and three-year-old infants and adults (Eiteljoerge et al., 2019b) supports this interpretation. Here, participants were familiarized with novel word-object and action-object mappings and then tested on their recognition of these mappings at the different ages. At test, they were shown the two novel objects side-by-side and then primed by hearing the label or seeing the action associated with one of the objects. Increased looking at the object which had previously been associated with either the prime label or action would indicate recognition and learning of the word-object or action-object mappings respectively. Looking patterns to the target object revealed that 1-year-olds learned action-object associations, but not word-object associations; 2-year-olds learned neither; and three-year-olds learned word-object associations, but not action-object associations; whereas college-aged adults showed learning of both. Similar findings are also reported by studies examining the simultaneous presentation of information from different modalities, finding that 15-month-olds, for instance, only learn word-object mappings when these were not accompanied by referential actions (Puccini & Liszkowski, 2012). Such findings are also in line with research showing that young children attend to input from a single preferred modality when presented with information from two modalities (e.g., Lewkowicz, 1994; Robinson & Sloutsky, 2004; Sloutsky & Napolitano, 2003).

Taken together, while the studies reported above suggest that there may be additional costs associated with learning from multisensory input, others suggest that multisensory input facilitates learning of novel word-object associations (Goldstein et al., 2020; Lewkowicz & Kraebel, 2004; Suarez-Rivera et al., 2019; Werker et al., 1998; Yu et al., 2006, 2008). The studies also point to developmental differences in the influence of multisensory environments on word-object learning, with very young infants profiting from them, while older infants may struggle with the complexity of such input, and older infants still showing few differences in action-object and word-object learning at later developmental stages (Childers & Tomasello, 2003; Dysart et al., 2016; Riggs et al., 2016). One potential reason for differences across these studies may be the timing of presentation of the stimuli, as we discuss next.

Timing in multisensory word-action-object learning

While some of the studies looking at multisensory environments assured that information from the linguistic and action modality were synchronously presented (i.e., 1- and 2-year-old's heard the words and saw the actions being performed on the objects simultaneously), others presented this information in a sequential manner, notably Eiteljoerge et al. (2019), (i.e., infants first saw the word-object mappings and then the action-object mappings or vice versa).

From a theoretical standpoint, the failure to find robust learning in such multisensory settings in the latter studies may be due to the sequential presentation of stimuli in their experiments. Gogate & Bahrick's (1998) theory of intersensory redundancy in early learning highlights the role of temporal alignment of vocalizations and motions and novel objects in early learning. For example, 7-month-olds learned and retained syllable-object relations only when stimuli were presented in synchrony with attention-grabbing movements, emphasizing that word-object mapping learning benefits from temporal alignment with actions. Mothers also used synchronous presentation of target words and object motion/touch more often than for non-target words (Gogate et al., 2000), and they exaggerated these motions when demonstrating them to their infants (Brand et al., 2002, 2007; Gogate, 2020). Older infants, however, may need to rely less on stimulus timing in word-object association tasks: mothers of pre-lexical infants used target words in synchrony with objects in motion more often than mothers of infants at later lexical stages. The authors conclude that mothers used synchronous labeling less often because their infants relied less on aligned stimulus presentation in order to map words onto novel objects (Gogate et al., 2000, 2001). Similarly, older infants are more likely to notice temporal misalignment of multimodal information presentation relative to younger infants (Gogate et al., 2014;

Lewkowicz, 1994; Matatyaho & Gogate, 2008), showing that sensitivity to temporal alignment increases in early development by facilitating attention allocation to the target object (Axelsson et al., 2012; Pomper & Saffran, 2019; Venker et al., 2018). Such findings are in keeping with the modality principle (Low & Sweller, 2005) according to which mixed mode presentations (partially visual, partially auditory) support learning processes more effectively relative to single mode presentation (either visual or auditory presentations; Low & Sweller, 2005). A similar explanation for effects of temporal synchrony is found in suggestions that the effect of speech is most prominent as infants uncover salient aspects of the present object in real-time, which may imply that mechanisms underlying multisensory information processing in early word learning tasks serve the unification of these information sources during encoding (Goldstein et al., 2020) by reducing the amount of modality-specific correspondences (Lany et al., 2020).

However, research on infant visual categorization with 1-year-olds suggests that synchronous presentation of labels and objects may negatively impact learning of word-object mappings. Althaus & Plunkett (2015) investigated learning of word-object associations when the objects were labeled in either a synchronous or sequential manner. Synchronous labelling of objects led to reduced preference for the target object at test relative to asynchronous labelling of objects, potentially due to higher processing costs during synchronous relative to asynchronous presentation of stimuli. Multisensory stimuli may lead to excessive cognitive load, due to working memory limitations on the representation and manipulation of information during complex cognitive tasks, before novel information is incorporated into long-term memory (Allen et al., 2006). Mentally integrating multiple sources of information can exceed working memory capacities (Sweller & Sweller, 2006), possibly hindering early learning of word-object and action-object mappings.

Thus, while there is considerable support for a positive influence of temporal systematicity in information presentation on learning, there are also recent studies suggesting the opposite. With a view to explaining differences in the effect of learning in multisensory environments, the current study will extend the findings of Eiteljoerge et al. (2019) by presenting young children with word-object-action mappings in a temporally synchronous and sequential manner and comparing learning across the two conditions, to examine whether impoverished learning in multimodal settings is influenced by the temporal aspects of information presentation.

Individual differences in multisensory word-action-object learning

Dynamic systems accounts of learning focus on the interaction between the child and her environment on learning, highlighting how differences in infant's attention to different aspects in their environment may drive differences in learning (Hirsh-Pasek et al., 2000; Karmiloff-Smith, 1997; Smith & Thelen, 2003; Thelen & Smith, 1994). Here, no specific aspect of the environment drives learning per se, but aspects may be more salient at different time points due to the interaction between the learners' abilities, interests and the complexity and shape of the environment. Given the emphasis on the learner as an additional source of variability in such theories of development (see also Gogate & Hollich, 2013 for a dynamic systems perspective on word learning), here we also examine the extent to which individual differences in language development (i.e., word comprehension and production) are linked to word-object and action-object mapping success at different time points.

Indeed, there is considerable research on the richness of individual variability in early word learning stages, with substantial variation in language comprehension and production over the developmental course. For instance, during the second year, some 18-month-olds may produce 50 words and more, while others only produce their first words by the end of the second year (e.g., Bloom, 2000; Brown, 1973; Clark, 2016). Such individual differences in vocabulary development may give rise to underlying causes of developmental change, forming the basis of theoretical underpinnings in early language acquisition (Bates et al., 1995; Frank et al., 2020). Here, exploratory analyses will examine the possibility of individual differences in language and action understanding on word-object and action-object learning in the current study. Specifically, we aim to identify developmental changes in word- and action-understanding based on the child's lexicon, allowing us to identify the extent to which variance in the data will be absorbed by language proficiency measures in the model. In what follows next, we describe the present study in more detail with regard to the key points highlighted above.

The current study

The current study investigated word-object and action-object associative learning from multisensory settings in one- and 2-year-olds. We used a similar design to Eiteljoerge et al. (2019) to examine both *replication* and *extension* of previous results.

As mentioned above, one potential reason why Eiteljoerge et al. (2019) and other studies failed to find evidence for robust learning at the ages tested may be the fact that the word and action information were presented in a temporally non-aligned manner in these studies, i.e., young children were first presented with the word-object mappings and then the action-object mappings or vice versa. Against the background that temporal systematicity in information presentation promotes learning in multimodal contexts (c.f., Gogate & Bahrick, 1998, 2001), we extended the findings of Eiteljoerge et al. (2019) by including a manipulation of temporal systematicity and presented 1- and 2-year-olds with the word-action-object mappings in either a temporally synchronous or sequential manner. Comparison of learning performance therefore allowed us to make stronger conclusions regarding learning in multisensory contexts across development, in particular with regards to the role of temporal systematicity in the presentation of linguistic and action information in learning in multisensory contexts at the ages tested.

Alternatively, it remains possible that infants and young children learned the mappings presented to them in Eiteljoerge et al. (2019) but nevertheless failed to show recognition of these mappings at test due to the complexity of the task. In particular, the preferential looking time task presented in Eiteljoerge et al., (2019) requires the child to match a word or an action to one of two images presented simultaneously on screen. Thus, the child must demonstrate knowledge of not only the fact that a particular word/action is associated with a particular object, but also that it is not associated with the distractor object. This may be difficult especially when learning is not as robust as was likely the case in this study (given the time course of target recognition reported across ages). In contrast, other tasks, such as simple habituation tasks or violation-of-expectation tasks merely require the child to notice a match or a mismatch between a previously trained association and the information currently presented (see also Ballem & Plunkett, 2005 for similar arguments, and 1.1. for further explanation), in the absence of a distractor image. This may be easier for the child to demonstrate, thereby allowing a more sensitive test of infant's learning of novel word-object and action-object associations, especially in cases where learning may not be as robust. Against this background, we tested learning of the word-object and action-object mappings using both the preferential looking paradigm used by Eiteljoerge et al., (2019) as well as an additional violation-of-expectation paradigm to obtain multiple indices of learning across a range of tasks with different levels of difficulty.

In particular, 1- and 2-year-olds were first trained on word- and action-object mappings as they were presented with two novel objects, each associated with a distinct novel name (e.g., "Oh, a Tanu!") and a distinct arbitrary action (e.g., a hand rotates the object around its axis). In a subsequent test phase, we examined 1- and 2-year-old's fixations to the target object (upon seeing the associated action being performed on a control object or hearing its associated label) as an index of their learning of the word- and action-object mappings. Half of the participants experienced words and actions during training in the presence of a novel object in a synchronous manner (i.e., input from linguistic and action modalities are presented simultaneously), and the other half in a sequential manner, (i.e., presentation of one modality proceeds the other). Comparison of performance across conditions allowed us to examine the impact of temporal systematicity in multisensory learning. In addition, we also included a separate violation of expectation task, where changes in 1- and 2-year-old's pupil dilation were be compared across word- and action-object match and mismatch trials. Increased pupillary dilation during mismatch trials relative to match trials indicate learning of the word-object and action-object associations (see Table 1). Recent findings suggests pupillometry to be a valid instrument when measuring infant learning and processing (Ackermann et al., 2019; Cheng et al., 2019; Hepach & Westermann, 2016; Pätzold & Liszkowski, 2019, 2020; Zhang & Emberson, 2020), as changes in pupil dilation diameter provide a momentary, involuntary and unbiased measure of arousal and cognitive load relative to looking time behavior (Beatty, 1982; Karatekin, 2004, 2007) and interpretation of pupillary data is relatively stable (e.g. no test order effects; Schöner & Thelen, 2006). Furthermore, we accounted for differences in 1- and 2-year-olds linguistic, and additionally in 1-year-olds cognitive and fine motor skills as potential factors influencing word-action-object associative learning.

Table 1

Illustration of Violation-of-Expectation Stimuli

Trial #	Training Stimuli			Match Trial Stimuli			Mismatch Trial Stimuli		
	Object	*Label*	*Action*	*Object*	*Label*	*Action*	*Object*	*Label*	*Action*
1	Green	Löki	pincing-rotating	Green	Löki	pincing-rotating	Green	Tanu	pincing-rotating
2	Green	Löki	pincing-rotating	Green	Löki	pincing-rotating	Green	Löki	grasping-inverting

Note. An example of match trial stimuli and mismatch trial stimuli in the context of the preceding training phase.

With regards to our examination of the effect of temporal systematicity (temporally synchronous, sequential) on learning (proportion of target looking at test), we predicted that the impact of temporal systematicity varies with potential benefits early on that wane across development, thus we expected to find an effect of temporal systematicity at 1 year but potentially not at 2 years of age. With regards to the influence of condition and in keeping with the literature highlighting the importance of actions early in development (Capirci et al., 2005; Caselli et al., 2012; Eiteljoerge et al., 2019b; Özçalışkan & Goldin-Meadow, 2005; Puccini & Liszkowski, 2012), we predicted learning of action-object associations to be more robust than word-object learning at 1 year of age. Furthermore, we predicted evidence for word-object learning at 2 years of age and expected stronger evidence for learning in the violation-of-expectation task relative to the preferential looking task. Finally, with regards to the exploratory analyses, we predicted that 1- and 2-year-olds who showed improved language skills will show stronger word-object learning relative to their peers.

2. Methods

2.1. Participants

This study was run on an entirely new sample of German-monolingual participants (i.e., infants hear German over 80% of the time in their daily life; Byers-Heinlein, 2015). 80 1-year-olds (44 females, M_{age} = 11.39 months, SD_{age} = 0.95 months, age range = 10 - 14 months) and 80 2-year-olds (40 females,

M_{age} = 23.03 months, SD_{age} = 1.102 months, age range = 22 - 26 months) participated in the study and data from additional infants were excluded due to calibration issues (*n* = 12) and fuzziness (*n* = 4; i.e., infant refused to be seated in front of the experimental screen and/or refused to fixate the pictures on the screen during the trials, and/or infant whined and cried consistently during the experimental procedure; Slaughter & Suddendorf, 2007;) during study participation.

Individual test trials were excluded if a participant looked at a stimulus for less than M – 3SD of the time (Eiteljoerge et al., 2019a). We anticipated a dropout rate of 26.1% overall based on the criteria used for data exclusion in Eiteljoerge et al., (2019) which was only 9.09% in the current study. Power analysis (using the *simr* package, Green & MacLeod, 2016) based on the generalized linear mixed models data (powering for a main effect of word-object and action-object association types) from 36-month-olds in Eiteljoerge et al. (2019) found that we had 90% power to detect the difference between conditions reported in this dataset (effect size = .54) with a sample size of 39 participants. Despite the fact that this study was over-powered, we tested 40 infants per condition, since effect sizes in younger populations may be smaller than in older populations. Indeed, we could not use the data from the 1-year-olds or 2-year-olds in the simulations since there was no difference between conditions at this age, although 1-year-olds showed learning of action-object associations.

1- and 2-year-olds were recruited from our Babylab database and their participation was rewarded with a book. Caregivers were asked to sign written informed consent for their child to participate in the experiment, as well as to complete the ELFRA-1 questionnaire (1-year-olds) or to fill in a short version (BabyLex Application, Mayor & Mani, 2019) of the MacArthur-Bates Communicative Development Inventories (2-year-olds; CDI, Fenson et al., 2007). The ethics committee of the XXX provided ethics approval for this project (No. 46/2018).

2.2. Stimuli

2.2.1. Preferential looking task

1- and 2-year-olds were presented with two novel objects (i.e., household appliances; a dough docker and a fruit press), two arbitrary actions and two novel names (Tanu and Löki). The labels were in keeping with German phonotactic constraints and were used in Eiteljoerge et al. (2019). Each object had two contrasting colors to present the child with easily distinguishable items (Gogate & Bahrick, 1998). The dough docker will henceforth be referred to as the "blue" object, while the fruit press will be the "green" object, see Figure 1.

Figure 9

Dough docker (blue) and fruit press (green) were modified and served as novel objects

2.2.1.1. Auditory stimuli

A female German native speaker produced the auditory stimuli in infant-directed speech. The auditory stimuli for the training phase consisted of blocked passages introducing the novel labels, with each passage presenting the respective novel label three times, i.e., "Ah! Ein Tanu! Oh, ein Tanu! Wow, ein Tanu!" (Ah! A Tanu! Oh, a Tanu! Wow, a Tanu!). The blocked presentation ensured that the child was presented with an adequate amount of repeated exposure to the label-object pairings and had provided robust word learning in the literature (Ackermann et al., 2019; Borovsky & Elman, 2006). The auditory stimuli for the test phase presented a single token of the target label, i.e., one of the novel labels, e.g., "Ein Tanu!" (A Tanu).

2.2.1.2. Visual stimuli

2.2.1.2.1. Training phase

We created four videos showing an agent's hand performing two distinct actions on both the novel objects, i.e., two videos per object. The videos began with the object standing in the middle of the screen and the agent's right hand to the right side of the object placed palm down on the table. Then the hand reached for the object and started to move the object according to the chosen actions.

In the pincer-rotating action presentation, the agent's right thumb and pointer finger reached for the center of the object and rotated the object once around its base and then placed the object back on the table (with the hand again by the side of the object facing down). This action was then repeated a further two times. In the grasping-reversing presentation, the agent's hand grasped the object and turned the object upside down and back up again (with the hand again by the side of the object facing down). This action was also then repeated twice. In all cases, the object was clearly visible to the infant when the action was performed on the object (see Fig. 1 for screenshots of the videos at critical times during the action performance). Each individual action was roughly 3s long. Training phase videos were 18 seconds long, 854 x 480 pixels in size and centered to the screen.

The auditory stimuli were merged with the videos described above separately for each condition. In the synchronous condition, the labels were always presented synchronously to the action movement, i.e., were presented during the main action. In the sequential condition, the labels were timed to occur when the right hand was lying flat on the table to the sides of the object. Thus, while the physical, auditory and visual stimuli used in the synchronous and sequential condition were identical, the timing of the audio and visual stimuli varied across the two conditions to allow either overlap of label-object and action-object pairings or separate non-overlapping presentation of label-object and action-object pairing. In the sequential condition, the presentation of action first or label first was counterbalanced across trials.

We chose these arbitrary actions over attention-getting (e.g., shaking) or goal-directed/meaningful actions to ensure that the labels and actions are comparable and due to 1- and 2-year-olds's potential familiarity with the attention-getting actions. This also ensured that the actions chosen in the current study were comparable to Eiteljoerge et al., (2019), while being distinct enough to allow for generalization of these results to different stimuli.

2.2.1.2.2. Test phase

The test phase was divided into a baseline window, prime window and recognition window. The visual images for the baseline and recognition window were identical and consisted of stationary images of the two objects presented side by side on screen for 2.5 seconds. The objects measured 240 by 410 pixels and were centered to the left and right side of the screen and were provided with no auditory stimuli during the baseline and recognition windows.

During the prime window, in action-object trials, infants saw a video of the agent's right hand performing one of the actions on a transparent Plexiglas tube (20cm long with a diameter of 50mm, PLEXIGLAS® XT), and in word-object trials, infants heard one of the labels directing their attention to one of the objects, i.e., the labelled target object, (e.g., "Ein Tanu!" (A Tanu!)). We used a transparent tube to ensure that the target object was never presented in the prime window (since we want to test 1- and 2-year-old's association of the action and the label with a habituated object). We chose to replace the object with a see-through, unambiguous object in preference to performing the action on an absent object (Eiteljoerge et al., 2019a) in order to avoid the possibility that empty hands performing an action may be confusing or ambiguous to the child.

In action-object trials, videos began with the Plexiglas tube standing in the middle of the screen and the agent's right hand to the right side of the object placed palm down on the table. Then the hand reached for the tube and started to move the object according to the chosen actions. In word-object trials, the prime window presented no visual stimuli, i.e., infants saw a blue fixation cross centered to a black screen as they heard the target label. The onset of the recognition window was exactly 200ms after the offset of the prime stimuli, i.e., following completion of the action (with the hand back on the table to the right side of the tube, palm facing down) or following completion of the target sentence. Presentation of visual prime stimuli (videos of action movement) was centered to the screen and lasted about four seconds.

2.2.2. Violation of Expectation (VOE) task

The objects, actions and labels we used in the Violation of Expectation paradigm (e.g., Baillargeon, 2004) were identical to those we used in the preferential looking paradigm. The objects, actions and labels were combined in two ways to create videos for match and mismatch trials in the VOE task. Match-trial-videos were either match-action-videos or match-label-videos. Match-action-videos presented 1- and 2-year-olds with one of the objects that they were familiarized with beforehand and saw the match action being performed on this object as in the training phase. Match-label-videos presented children with one of the objects that they were familiarized with beforehand as they heard the label for this object from the training phase. Mismatch-trial-videos were either mismatch-action-videos or mismatch-label-videos. Mismatch-action-videos presented 1- and 2-year-olds with one of the objects that they were familiarized with beforehand as they saw a different action (i.e., the other action) being performed on this object than the one they were presented with in the training phase. Mismatch-label-videos presented children with one of the objects that they were familiarized with beforehand as they heard a different label (i.e., the other label) for this object relative to the one from the training phase. Thus, when the green object was associated with the label Löki and was rotated in the training phase, in mismatch trials, the green object was presented with the label Tanu for half of the trials and was inverted for the other half, see Table 2.

Table 2

Experimental Design

Phase	Number of trials	Example
Training	4	Each object will be presented twice (with distinct label and action shown three times each)
Preferential looking - test phase	8	Baseline - Prime - Recognition (four word-object & four action-object trials; 4 trials per object where this object is the target)
Violation of expectation - test phase	8	Each object will be presented 4 times. Once with the associated label from the training phase, once with the associated action from the training phase, once with the other label from the training phase and once with the other action from the training phase.

Note. Preferential looking paradigm including training and test phase (see also Eiteljoerge et al., 2019), followed by a VOE trial sequence.

2.3. Language proficiency

In addition, we collected vocabulary data using the ELFRA-1 questionnaire for 1-year-olds (Parent questionnaire for 1-year-olds: Language, gestures and fine-motor skills, Grimm & Doil, 2000) and the short adaptation “BabyLex” (Mayor & Mani, 2019) of the Mac-Arthur Bates Communicative scale for the 2-year-olds. ELFRA-1 comprised a list of 164 words where parents could indicate whether their child “understands/produces” or “doesn´t understand/produce” a critical word. They also responded to 67 questions (“yes” or “no” answers) on their child´s development across four scales (language production, language comprehension, gestures, fine motor skills). BabyLex robustly assesses 1- and 2-year-old´s vocabulary knowledge in short administration time by estimating the full CDI-score based on parental responses on a set of 25 words.

2.4. Procedure

The child sat either in a baby seat or on their caregiver´s lap at a distance of 65cm away from a TV screen (92 x 50cm) in a dimly lit experimental room. A remote eye tracker (Tobii X 120) placed underneath the TV screen recorded gaze data at 60 Hz and Tobii Pro Labs software was used for stimulus presentation. Two loudspeakers placed above the TV screen presented the auditory stimuli, while the visual stimuli were presented on the TV screen. Two cameras located immediately above the TV screen provided online recordings of the child and were used to keep track of the child during the course of the experiment. We used a five-point grid calibration (a dot appearing in every corner and one in the center of the screen) in Tobii Studio. The experiment was initiated following successful calibration and each trial began with an attention getter to reorient 1- and 2-year-old’s attention to the screen and the trial only started when 1- and 2-year-olds were fixating the screen. The experimenter initiated the beginning of a new trial via key press from the host computer only when the child attentively fixated the screen. The eye-tracking experiment took about 6 minutes.

2.5. Experimental Design

The main experiment consisted of a preferential looking task where 1- and 2-year-olds were introduced to novel action-word-object triads and tested on their learning and recognition of these novel word-object and action-object pairings, as well as a VOE task which presented 1- and 2-year-olds with match trials, where they saw the same word-object and action-object pairings that they had seen during training, and mismatch trials, where 1- and 2-year-olds saw the object associated with the other action or other word. This allowed us to investigate their pupillary responses to violations of recently trained word-object and action-object associations (see Table 1).

2.5.1 Preferential Looking Paradigm

2.5.1.1. Training Phase

The training phase consisted of 4 trials. In each trial, 1- and 2-year-olds saw a distinct action being performed on one of the novel objects while hearing the label for this object. Each trial presented three word-object and three action-object pairings, i.e., 1- and 2-year-olds saw a single novel object on screen as the action was repeatedly performed on this object three times and the label for this object was presented three times. 1- and 2-year-olds saw two trials each for each novel object, i.e., in total they saw the action associated with each object six times and hear the label for each object six times as well. This presented 1- and 2-year-olds with considerably more tokens of the novel label than in Eiteljoerge et al., (2019; who presented infants and young children with only one repetition of the input) as well as varying the clustering of tokens since such clustered presentation of mappings has shown robust learning in previous studies (Ackermann et al., 2019; Borovsky & Elman, 2006).

To test the effect of temporal systematicity in word-object and action-object associative learning, half of the 1- and 2-year-olds heard word-object and saw action-object presentations in a temporally synchronous manner and the other half in a temporally sequential manner. In the sequential condition, word and action presentations were sequential; e.g., a child first saw the action performed on the object once, and then heard the label for this object as soon as the action performed on the object is completed, followed by a subsequent sequential presentation of action-object and word-object pairings. Action-object and word-object presentations lasted about three seconds each and were repeated six times for each object (videos lasted roughly 18 seconds). Order of stimulus presentation was counterbalanced within participants. Thus, in one trial, 1- and 2-year-olds saw the action-object pairing first while seeing the word-object pairing first in the other trial. During action-object presentations, the child saw the object centered to the screen with the agent´s hand performing an arbitrary action on the object. For word-object presentations, the child heard the label for the object as they saw the object centered to the screen with the agent´s right hand on the table by the side of the object, palm facing down.

In the synchronous condition, onset of word-object and action-object presentations was aligned and lasted about three seconds. Thus, 1- and 2-year-olds heard the object being labeled while seeing the action being performed on the object. Within each trial, temporally aligned word-object-action presentations were repeated three times for each object and 1- and 2-year-olds were presented with two trials per object. To match video lengths across conditions (18 seconds), for the synchronous condition, each stimulus presentation was followed by an interval of three seconds where the child saw the object centered to the screen with the agent´s hand on the table by the side of the object, palm facing down (complementing word-object presentation in the sequential condition). During this inter-stimulus-interval, no label or action information will be presented to the child.

Across participants, we counterbalanced the distinct actions that were associated with each object as well as the labels associated with each object, such that half of the 1- and 2-year-olds saw the blue object being rotated and/or labeled Löki and the other half of the 1- and 2-year-olds saw the same object being turned upside down and/or labeled Tanu. Thus, four training lists were used in the experiment with the trial order being fully randomized across training and test phase, and 1- and 2-year-olds were randomly assigned to the synchronous and sequential condition.

2.5.1.2. Test Phase

1- and 2-year-olds were presented with 8 test trials, where 4 trials examined infant's recognition of the trained action-object associations (two per object) and 4 trials examined infant's recognition of the word-object associations (two per object). At test, each trial consisted of a baseline phase, prime phase and recognition phase. Baseline and recognition phase were identical and involved the silent presentation of stationary images of the two objects centered to the left and to the right of the screen for 2.5 seconds each (Von Holzen & Mani, 2012). The location of the object on screen was counterbalanced across trials.

Each trial began with the baseline phase, followed by the prime phase. During the prime phase (about 4 seconds long), infants either heard the label for one of the objects (word-object trial) or saw one of the arbitrary actions being performed on a transparent, unambiguous control object (action-object trial). In word-object trials, infants received no visual stimuli during presentation of the object label, and in action-object trials, infants received no auditory stimuli during presentation of the object action. The recognition phase began 300ms after the offset of the label or the action and the location of the two objects onscreen was identical across baseline and recognition phases.

2.5.2. VOE Paradigm

2.5.2.1. Test Phase

The test phase included eight trials, with four match trials and four mismatch trials (two match and two mismatch trials per object). In *match word trials* (one per object), the child saw a stationary image of one of the objects (e.g., Object A) and heard the label that was presented with Object A in the training phase. In *match action trials* (one per object), the child saw the match action being performed on Object A that they will have seen in the training phase. During *mismatch word trials* (one per object), the child saw Object A and heard the label that was associated with Object B in the training phase. In *mismatch action trials* (one per object), the child saw the action that was performed on Object B (in the training phase) now being performed on Object A. Each trial began with a 1-s baseline phase where 1- and 2-year-olds saw a stationary image of the object on screen in silence, followed by the onset of the language or action information. Each trial ended 1 second after the offset of this information. The order of presentation of trials was counterbalanced within and across participants.

2.6. Pre-processing

2.6.1. Preferential Looking Paradigm

Gaze data was collected using Tobii X 120 eyetracker (60Hz) which provided data for where a child is looking every 16.7 ms in a trial. Tobii Pro Labs data included gaze data information and trial data information which was further processed and analyzed in R (R Core Team, 20). Data from a time stamp was only included when one or both eyes were tracked. For time stamps where reliable data were provided for both eyes, we computed mean gaze points averaged across both eyes. For time stamps where reliable data were provided only for one eye, we used the data from only this eye. Gaze data was then aggregated into 40ms bins.

Two areas of interest were defined according to the location and size of the stationary images of the objects presented on the screen during the baseline and the recognition phase (test phase). For each time point, we automatically calculated whether the child looked at the target object, the distractor object, or at neither of these on the screen. To ensure that we only considered eye movements that could reasonably be interpreted as a response to the information presented in the prime phase, we only included eye movements that began 240ms and 2000ms after the onset of the recognition phase (cf. Eiteljoerge et al., 2019), and related these to eye movements in a corresponding time window from 240ms to 2000ms from the onset of the baseline phase. We calculated for each phase the proportion of target looking separately, which represented the duration of time that 1- and 2-year-olds spent looking at the target image (i.e., the image whose associated action or label had been presented during the prime phase) over the duration of time that they spent looking at the target and distractor image. For the generalized linear mixed models (henceforth GLMM, described in detail below), we calculated this measure separately for each time stamp in the recognition phase. We then corrected the proportion of target looking in the baseline phase, i.e., we subtracted the average proportion of target looking in the baseline phase from the average proportion of target looking in the recognition phase, or from each time stamp in the recognition phase (for the GLMM). This ensures that fixations in the recognition phase were relativized with respect to the baseline phase. Further analyses were then performed based on this baseline-corrected fixation score (dependent variable).

Test trials were excluded from further analyses when participants looked at the stimuli for less than *M* - 3*SD* of the time, since short fixations and rapid eye-movements between objects are unlikely to be reliable indicators of preferential target looking. Participants were excluded from the analyses if they did not provide data for any one of the conditions during the test phase.

2.6.2. Violation of Expectation Paradigm

We also examined 1- and 2-year-old's pupillary responses to violations of the associations from the training phase. Here, we only included timestamps in the analysis that provided valid eye tracking data, using the validity criterion output by the eye-tracker. Then we calculated for each eye separately the difference in pupil size between two consecutive data points (see also Ackermann et al., 2019). Data points for which the difference between these two neighboring data points rank in the top or bottom 10% of overall differences between neighboring data points were excluded. Thus, large deviations between neighboring data points - which were likely to be artefacts in the data - were excluded from further analyses. Missing data were accounted for by data interpolation across a moving window of four consecutive data points (Hepach et al., 2012). Finally, we averaged the filtered, interpolated data across both eyes to provide a measure of pupillary size for each bin.

Participants who provided data for less than three test trials were excluded from the study to ensure that valid data averaged across a suitable number of trials was included in the analyses. Second, participants were also excluded if they only provided data for one of the two conditions in the test phase. This means that each child needed to provide data for at least one trial per condition at test providing data in both the baseline and recognition phase in two trials in at least one condition. Third, trials where participants looked at the stimuli for less than M - 3SD of the time in a single test trial were excluded from further analyses.

We binned the data into 40ms bins to ease interpretation of time-sensitive changes in the pupil data. The baseline phase represents the first part (i.e., pre-stimulus) of the trial in which only the object is shown and which is followed by the test phase, initiated with stimulus onset (i.e., action or label, while still showing the object). We calculated a joint baseline mean for both pupils for each pre-stimulus phase which was then subtracted from the pupil mean for each time bin during the test phase, resulting in a baseline-corrected pupil measure.

3. Analysis

To avoid complicated interactions with age which may be difficult to interpret, all analyses were performed separately in each age group. For both age groups, we used identical formulations for the full and null models, fixed effects, random effect structures and the dependent variables (baseline-corrected PTL, changes in pupil diameter), the same software version (version 4.0.5; R Core Team, 2020) to preprocess, model and plot the data, as well as identical functions and their corresponding packages (and package versions) to analyze the data.

3.1. Preferential looking task

3.1.1. Word-object and action-object learning and their temporally aligned presentation

First, to estimate the effects of *condition* (word-object, action-object) and *temporal systematicity* (synchronous, sequential presentation) on 1- and 2-year-old’s baseline-corrected PTL, we used a

GLMM (Baayen et al., 2008) and included these two within-participant factors and their interaction as fixed effects. That is, we also controlled for a potential interaction of the two effects, such that synchrony could have different effects in word-object and action-object learning, see model specification: Model 1.

Model 1 = lmer (PTL ~ Condition * Temporal systematicity + (Condition | ID) + (1 | Object), Data = data, REML = F, control = contr).

We fitted the data using the function lmer of the package lme4 (version 1.1-23; Bates et al., 2015). Since we have repeated observations of the same individuals, we included *participant ID* as a random intercepts effect, and *condition* within *participant ID* as a random slope (to assure type I error rate at the nominal level of 0.05; Barr, 2013; Schielzeth & Forstmeier, 2009). We also added *object* as a random intercept effect, whereby the term *object* represents the prime that was associated with the target object at test (i.e., four levels, two actions and two labels). As an overall test of the fixed effects of *condition* and *temporal systematicity* and their interaction and to avoid multiple cryptical testing, we conducted a full-null model comparison (Forstmeier & Schielzeth, 2011) based on a likelihood ratio test (Dobson, 2002) with the null-model being identical to the full model except for lacking the fixed effects of *condition* and *temporal systematicity*.

Second, to account for changes in infant looking behavior over the course of the trial within the model, we examined the effect of *condition* and *temporal systematicity* interacting with the factor *time* and its first, second and third polynomial (Mirman, 2016; Mirman et al., 2008, see also Eiteljoerge et al., 2019). We again fitted a GLMM including *condition*, *temporal systematicity*, and the three orthogonal polynomials and their interactions as fixed effects and again used the function lmer of the package lme4 (version 1.1-23; Bates et al. 2015, see model 2 specification). We again included *participant ID* and *object* as random intercepts, and the first, second and third orthogonal polynomial within *participant ID* and *object* as random intercept and slope (again, to keep type I error rate at the nominal level of 0.05). Again, the full-null model comparison was run with the null-model lacking the two effects of interest – *condition* and *temporal systematicity* – but still containing the three orthogonal polynomials to represent time over the course of the experiment and being identical with regard to the random effect structure of the full model.

Model 2 = lmer (PTL ~ Condition * Temporal systematicity * (ot1 + ot2 + ot3) + (1 + ot1 + ot2 + ot3 | ID) + (1 + ot1 + ot2 + ot3 | Object), Data = data, REML = F, control = contr).

After fitting the models, we checked whether the assumptions for normally distributed and homogeneous residuals were fulfilled by visual inspection of a QQ-plot of residuals (Field, 2005) and residuals plotted against fitted values (Quinn & Keough, 2002). As in Eiteljoerge et al., (2019), drop 1 analyses circled through the fixed effects to examine the influence of adding *condition* and *temporal systematicity* as predictors to the model, to examine best fit using a chi-squared likelihood ratio test. Effects of either predictor were then examined separately as motivated by the results of Model 1, i.e., pivoting on significant predictors. To obtain confidence intervals for the final reduced model estimates we used a parametric bootstrap (function bootMer of the package lme4; *N*=1.000 bootstraps).

3.1.2. Language proficiency

To examine the extent to which 1- and 2-year-old's language and 1-year-old's gesture and fine-motor development predicts variance for each type of learning (word-learning and action-learning), developmental estimates entered the model as fixed effects. The dependent variable was again baseline-corrected PTL. We split the data by *condition* and fitted a GLMM (Baayen et al., 2008) with the function lmer of the package lme4 (version 1.1-23; Bates et al., 2015) for each of the two conditions. Because there were two observations per condition for each participant (e.g., one label for each of the two objects such that every child was presented with two different labels), as well as to minimize type I error rate, we added *participant ID* as a random intercept to the model. To identify the potential contribution of individual predictors in the fixed effects structure, we used drop1 analyses (stats package version 4.0) explore the extent to which individual changes in infants looking time patterns could be explained by the variance of language abilities at 1 or 2 years of age and/or fine motor skills at 1 year of age.

Model_1-year = lmer(bc_PTL ~ ELFRAcompr + ELFRAprod + ELFRAgestures + ELFRAfmot + (1|id), data = (word_data or action_data), REML = F, control = contr)

Model_2-years = lmer(bc_PTL ~ Bperc + (1|id), data = (word_data or action_data), REML = F, control = contr)

3.2. Violation of Expectation task

We fitted a GLMM using the function lmer of the package lme4 (version 1.1-23; Bates et al., 2015). In addition to *condition* (action or word with object), *temporal systematicity* (synchronous or sequential presentation) and *VOE trial type* (mismatch/match) as well as all of their interactions entered the model as fixed effects. We included *object* (here, object represents the associative links between object, action and word presented during training, i.e., eight levels) and *participant ID* as random intercepts effect to account for repeated observations per infant (eight test trials), and added *condition* and *match* within *participant ID* as random slope. We conducted a full-null model comparison with the null model being identical to the full model, except for lacking the critical 3-way-interaction (*condition* interacting with *match* and *temporal systematicity*).

Model 3 = lmer (PD ~ Condition*Match*Temporal systematicity + (Match| ID) + (Condition | ID) + (1 | Object), Data = data, REML=F, control = contr).

Drop 1 analyses circled through the fixed effects to examine the influence of adding *condition*, *match* and *temporal systematicity* as predictors to the full model, to examine best fit using a chi-squared likelihood ratio test. Effects of each predictor was then examined separately as motivated by the results of Model 3, i.e., pivoting on significant predictors.

4. Results

4.1. 1-year-olds

4.1.1. Preferential looking task

The final sample provided 315 trials from which 19 trials were excluded (which removed data from two infants) because participants looked at stimuli for less than *M* - 3SD at a time, leaving 296 trials (92.5% of all trials shown) from 78 infants for further analyses (n_{action} = 149 trials, n_{word} = 147 trials).

Table 3

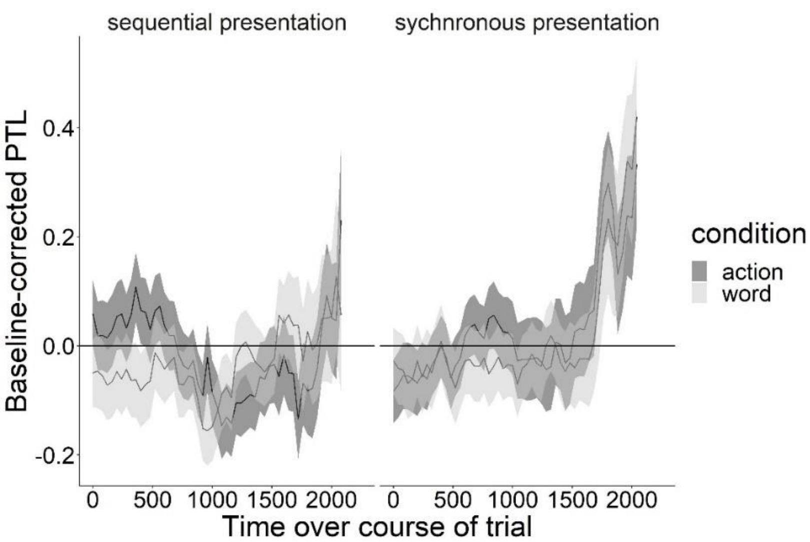
1-year-olds: GLMM estimating effects of condition, temporal systematicity, and their interaction on proportional target looking (PTL)

	baseline-corrected PTL at 1 year				
Predictors	*Estimates*	*std. Error*	*CI*	*Statistic*	*p*
(Intercept)	-0.017	0.037	-0.089 – 0.054	-0.476	0.634
condition [word]	-0.043	0.052	-0.145 – 0.058	-0.834	0.404
Temporal systematicity [sync]	0.005	0.052	-0.098 – 0.107	0.093	0.926
condition [word] * Temporal systematicity [sync]	0.043	0.074	-0.102 – 0.189	0.581	0.561
Random Effects					
σ^2	0.10				
$\tau_{00\ id}$	0.00				
$\tau_{00\ prime.used}$	0.00				
$\tau_{11\ id.condition.code}$	0.00				
$\rho_{01\ id}$					
N_{id}	78				
$N_{prime.used}$	4				
Observations	296				

Visual data inspection showed that assumptions for normally distributed and homogeneous residuals were met. There were some expected deviations of homogeneity assumptions for residuals plotted against fitted values which are rather common for looking time data. The full - null model comparison was not significant, $\chi 2(3) = 1.194$, $p = .754$, and testing individual fixed effects using drop1 (stats package version 4.0; i.e., comparing the overall model including all input variables with the model resulting from removing one of these variables) revealed no significant impact of *condition* interacting with *temporal systematicity* on infant's baseline-corrected PTL ($p = .56$), see Table 3 and Figure 10[1]. We estimated model stability by dropping data from individual infants one at a time from the full data set, comparing the model estimates derived from these subsets with the estimates for the full data set. This revealed the model to be of good stability (see Table A2).

Figure 10

1-year-olds: Changes in baseline-corrected proportional target looking (PTL) as a function of condition (word vs. action) and temporal systematicity (sequential vs. synchronous)



Next, we examined the data across the course of the trial by including *time* and its first, second and third polynomial. Due to the complexity of interactions included in this model, the steps we took to reduce this model in order to ease interpretability, and the similarity in the results of the models including and not including *time*, we report these results in the supplementary information. In short, we did not

[1] We also report an updated version of the registered model with adapted random effect structures including the intercept and slope of *condition* (dummy coded & centered) within *participant ID* and intercept and slope of *synchronicity* (dummy coded & centered) within primes shown at test (i.e., four levels, two actions and two labels). Log likelihood ratio tests of the updated model ($\chi 2(11) = -80.209$) and a simpler version of the updated model (i.e., lacking intercept and slope correlations in the random effect structure and estimating all random effects to be 0; $\chi 2(9) = -80.209$) revealed similar likelihood estimates such that the simpler model of the two was used for further analyses. Estimates for the critical interaction in the updated model and the registered model reported here were identical (p = 0.561), see Table A1.

find robust evidence for either learning of the word-object associations or action-object associations at this age, nor did we find any effects of either *condition* or *temporal systematicity*.

Finally, we provide an overview of infants' mean baseline-corrected PTL in each *temporal systematicity* group and per *condition*, as well as group differences contrasted against 0 (i.e., chance looking), see Table 4.

Table 4

1-year-olds' baseline-corrected PTL in each temporal systematicity group and per condition

Temporal systematicity	*Condition*	*M*	*SD*	*t*	*p*	*df*	*lower.CI*	*upper.CI*
Synchronous	action	-0.01	0.33	-0.33	0.74	72	-0.09	0.06
	word	-0.01	0.32	-0.34	0.74	71	-0.09	0.06
Sequential	action	-0.02	0.32	-0.48	0.63	75	-0.09	0.05
	word	-0.06	0.32	-1.66	0.10	74	-0.13	0.01

Note. Descriptives of baseline-corrected PTL in the recognition phase of the synchronous and sequential group and per condition. Values greater than 0 indicate looks to target and values smaller than 0 indicate distractor looks. Additional t-test estimates describe whether group means significantly differ from chance looking indicated as 0, i.e., similar proportions of looks to target and distractor.

4.1.2. Language proficiency

Parental reports on infants' word production and comprehension, as well as their use of gestures and fine motor skills served to explore the extent to which such early developmental milestones predict infant's object recognition based on word and action cues. In the word-learning data, correlations among the four predictors were low, suggesting that there is no evidence of multicollinearity following the "rule of thumb" for VIF > 5 (Chatterjee & Simonoff, 2013) for word production (*VIF* = 1.221), word comprehension (*VIF* = 1.512), gesture use (*VIF* = 2.072), and fine motor skills (*VIF* = 1.779). Visual inspection of the data revealed normally distributed data points and assumptions about homogeneous residuals and residuals plotted against fitted values were met. Drop1 analyses (stats package version 4.0) suggested that individual changes in infants looking time patterns could not be explained by variance of any of the predictors, see Table 5. Model estimates were of good stability, see Table A11.

Table 5

1-year-olds, word-learning: Individual differences in baseline-corrected PTL

	baseline-corrected PTL at 1 year				
Predictors	*Estimates*	*std. Error*	*CI*	*Statistic*	*p*
(Intercept)	-0.075	0.113	-0.297 – 0.146	-0.665	0.506
ELFRAcompr	0.001	0.002	-0.002 – 0.004	0.372	0.710
ELFRAprod	0.009	0.006	-0.003 – 0.020	1.496	0.135
ELFRAgestures	-0.001	0.008	-0.016 – 0.015	-0.084	0.933
ELFRAfmot	-0.006	0.017	-0.039 – 0.027	-0.359	0.720
Random Effects					
σ^2	0.10				
$\tau_{00\ id}$	0.00				
N_{id}	73				
Observations	139				

As for the action-learning data, multicollinearity among the predictors was again low, i.e., for word production (*VIF* = 1.202), word comprehension (*VIF* = 1.534), gesture use (*VIF* = 2.151), and fine motor skills (*VIF* = 1.809), and distribution patterns and assumptions of homogeneity were similar to the word-learning model. Drop1 analyses (stats package version 4.0) revealed that none of the predictors explained individual changes in infant's target looking (just like for the word-learning data) see Table 6. Model estimate ranges were of good stability, see Table A12.

Table 6

1-year-olds, action-learning: Individual differences in baseline-corrected PTL

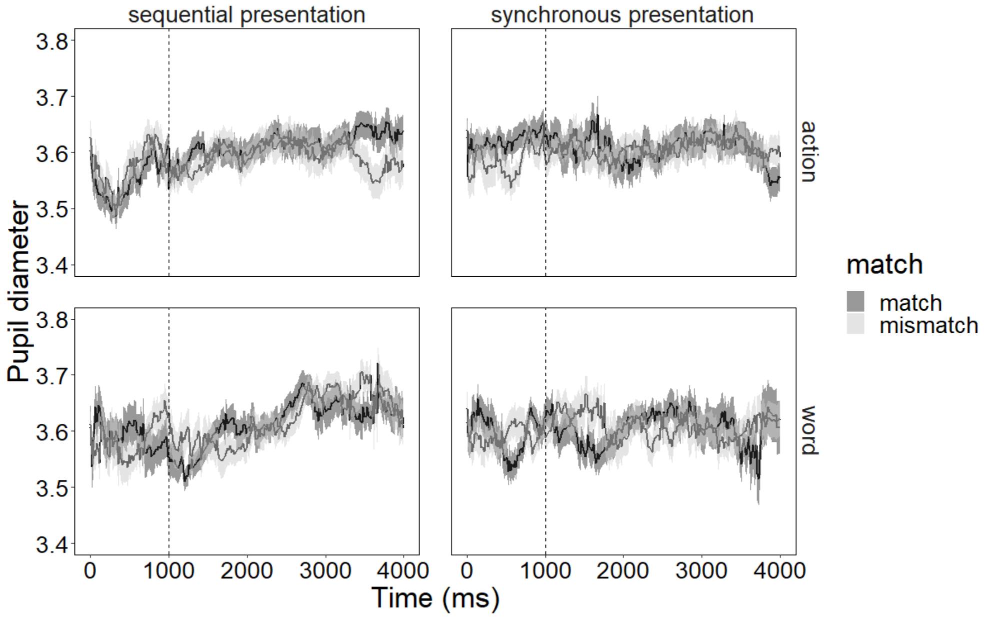
	baseline-corrected PTL at 1 year				
Predictors	*Estimates*	*std. Error*	*CI*	*Statistic*	*p*
(Intercept)	-0.075	0.113	-0.297 – 0.146	-0.665	0.506
ELFRAcompr	0.001	0.002	-0.002 – 0.004	0.372	0.710
ELFRAprod	0.009	0.006	-0.003 – 0.020	1.496	0.135
ELFRAgestures	-0.001	0.008	-0.016 – 0.015	-0.084	0.933
ELFRAfmot	-0.006	0.017	-0.039 – 0.027	-0.359	0.720
Random Effects					
σ^2	0.10				
$\tau_{00\ id}$	0.00				
N_{id}	73				
Observations	139				

4.1.3. Violation of Expectation task

All 80 infants delivered pupil data for at least three trials in the VOE task, providing a total of 620 trials (96.9% of all trials shown) after the top and bottom ten percent of overall differences between two adjacent samples (i.e., to avoid large deviations between neighboring data points) and invalid datapoints (as defined by the Tobii X120 eye tracker system) had been removed. We excluded 0.1 % of the pupil data where infants looked at stimuli for less than M - 3SD at a time and removed another 4.8% of missing data that could not be accounted for by data interpolation across a moving window of four consecutive data points from the full data set, leaving us with 610 trials (95.31% of all trials shown) from 80 infants to enter the model (n_{action} = 312 trials, n_{word} = 298 trials). Changes in infants' baseline-corrected pupil dilation in each group are depicted in Figure 11.

Figure 11

1-year-olds: Changes in baseline-corrected pupil dilation for every combination of condition (word vs. action) and temporal systematicity (sequential and synchronous) in match and mismatch trials



Note. The dotted line indicates onset of label or action.

The pupillary data were normally distributed and assumptions about homogeneous residuals and residuals plotted against fitted values were met. A model comparison revealed that the full model explained infant pupillary changes significantly better than the null-model ($\chi 2(7) = 45.721$, $p < .001$), however, drop1 analyses (stats package version 4.0) did not signal contribution of the interaction term on the dependent variable ($p = 0.421$, see Table 7). The model was of good stability (i.e., estimates derived from the model were in close range to the estimates derived from model subsets as we dropped data from individual infants one at a time), see Table A13.

Table 7

1-year-olds: GLMM estimating interaction effects of condition, temporal systematicity and match on pupil diameter changes

	baseline-corrected pupil dilation at 1 year				
Predictors	*Estimates*	*std. Error*	*CI*	*Statistic*	*p*
(Intercept)	0.557	0.038	0.483 – 0.632	14.659	**<0.001**
condition [word]	-0.014	0.036	-0.084 – 0.056	-0.401	0.688
match [mismatch]	0.021	0.029	-0.035 – 0.078	0.735	0.462
synchronicity [sync]	0.006	0.049	-0.090 – 0.103	0.131	0.896
condition [word] * match [mismatch]	-0.003	0.039	-0.080 – 0.073	-0.084	0.933
condition [word] * synchronicity [sync]	0.008	0.040	-0.070 – 0.087	0.207	0.836
match [mismatch] * synchronicity [sync]	0.007	0.041	-0.073 – 0.087	0.177	0.860
(condition [word] * match [mismatch]) * synchronicity [sync]	-0.026	0.055	-0.134 – 0.081	-0.482	0.630
Random Effects					
σ^2	0.03				
$\tau_{00\ id}$	0.02				
$\tau_{00\ id.1}$	0.01				
$\tau_{00\ item.x}$	0.00				
$\tau_{11\ id.condition.code}$	0.00				
$\tau_{11\ id.1.matchmismatch}$	0.00				
$\rho_{01\ id}$	-1.00				
$\rho_{01\ id.1}$	-1.00				
N_{id}	80				
$N_{item.x}$	8				
Observations	610				

4.2. 2-year-olds

4.2.1. Preferential looking task

We obtained a total of 293 trials (91.56% of all trials shown) in the preferential looking task (n_{action} = 144, n_{word} = 149). The data were normally distributed and assumptions about homogeneous residuals were met with some expected deviations of homogeneity for residuals plotted against fitted values. The full-null model comparison revealed that the full model explained 1- and 2-year-olds's looking behavior at test significantly better than the null-model lacking *condition* and *temporal systematicity* ($\chi2(3) = 15.495$, $p < .001$). Drop1 analyses (stats package version 4.0) were used to test the effect of individual fixed effects by comparing the full model to models resulting from removing one of the predictors and found no significant contribution from the interaction of *condition* and *temporal systematicity* on infant's baseline-corrected PTL ($t = -1.526$, $SE = 0.075$, $p = 0.127$), see Table 8. Model stability was estimated by comparing model estimates of the full model to those derived from model subsets (i.e., by dropping data from individual participants one at a time) and revealed good stability (see Table B2).

Table 8

2-year-olds: GLMM estimating effects of condition, temporal systematicity and their interactions on proportional target looking (PTL)

	baseline-corrected PTL at 2 years				
Predictors	*Estimates*	*std. Error*	*CI*	*Statistic*	*p*
(Intercept)	-0.091	0.035	-0.158 – -0.023	-2.614	**0.009**
condition [word]	0.105	0.054	-0.001 – 0.210	1.942	0.052
Temporal systematicity [sync]	0.177	0.049	0.081 – 0.273	3.611	**<0.001**
condition [word] * Temporal systematicity [sync]	-0.115	0.076	-0.263 – 0.033	-1.527	0.127
Random Effects					
σ^2	0.09				
$\tau_{00\ id}$	0.00				
$\tau_{00\ prime.used}$	0.00				
$\tau_{11\ id.condition.code}$	0.02				
$\rho_{01\ id}$	1.00				
N_{id}	80				
$N_{prime.used}$	4				
Observations	293				

Because the full-null-model comparison showed a significant contribution of the predictors in the interaction term (i.e., the interaction itself was not significant, p = .127), we reduced the full model by taking out the interaction to investigate the potential contribution of the individual predictors instead. This reduced model showed that the *temporal systematicity* of stimulus presentation, i.e., whether the words and actions were presented synchronously or sequentially, significantly predicted 2-year-old's target looking patterns at test, t = 3.48, SE = 0.036, p < .001 (see Figure 12), while *condition*, i.e., whether we examined 2-year-old's learning of action-object or word-object associations, did not (p = .227), see Figure 12. 2-year-olds' mean baseline-corrected PTL in each *temporal systematicity* group and per *condition*, as well as group differences contrasted against 0 are depicted in Table 9.

Table 9

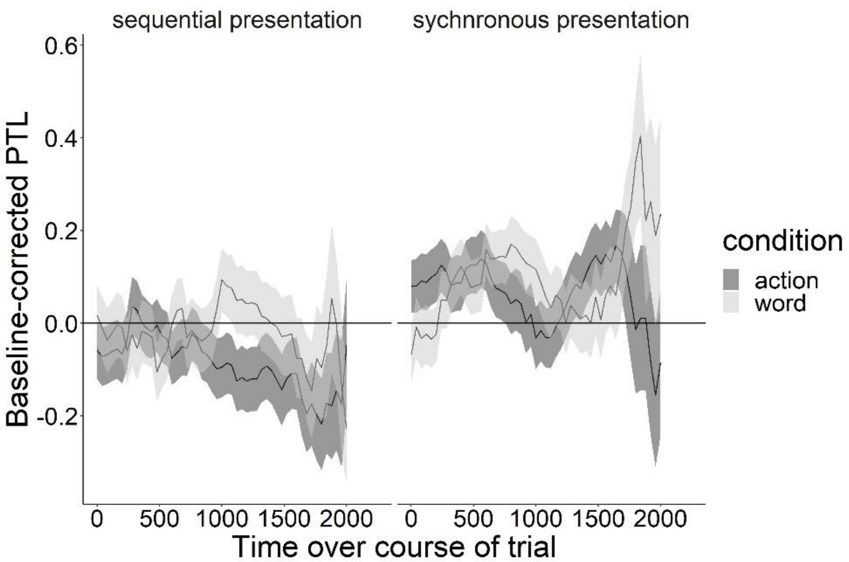
2-year-olds: baseline-corrected PTL in each temporal systematicity group and per condition

Temporal systematicity	*Condition*	*M*	*SD*	*t*	*p*	*df*	*lower.CI*	*upper.CI*
synchronous	action	0.087	0.292	2.512	0.014	72	0.017	0.154
	word	0.075	0.347	1.891	0.062	76	-0.004	0.153
sequential	action	-0.090	0.261	-2.941	0.004	71	-0.151	-0.029
	word	0.015	0.328	0.379	0.705	71	-0.063	0.092

Note. Descriptives of baseline-corrected PTL in the recognition phase of the synchronous and sequential group and per condition. Values greater than 0 indicate looks to target and values smaller than 0 indicate distractor looks. Additional t-test estimates describe whether group means significantly differ from chance looking indicated as 0, i.e., similar proportions of looks to target and distractor.

Figure 12

2-year-olds: Changes in baseline-corrected PTL as a function of condition (word vs. action) and temporal systematicity (sequential vs. synchronous)



Next, we examined the effect of *condition* and *temporal systematicity* interacting with the factor *time* and its first, second and third polynomial within the model to examine 2-year-old’s looking time patterns over the course of the test trials. As with the 1-year-olds, there was considerable overlap between the results of the model including and excluding *time*. We, therefore, report these analyses in the supplementary information in order to simplify presentation of the results.

4.2.2. Language proficiency

Visual inspection of the data revealed normally distributed data points, and homogeneity assumptions were met for both models. The two models revealed that parental reports on 2-year-old's word production did not predict word-object (see Table 10) or action-object (see Table 11) recognition at test, and both models were of good stability (see Table B9 and B10).

Table 10

2-year-olds, word-learning: Individual changes in baseline-corrected PTL

	baseline-corrected PTL at 2 years				
Predictors	*Estimates*	*std. Error*	*CI*	*Statistic*	*p*
(Intercept)	0.033	0.069	-0.102 – 0.168	0.475	0.635
BabyLex.percentage.score	0.000	0.001	-0.002 – 0.002	0.168	0.867
Random Effects					
σ^2	0.11				
$\tau_{00\ id}$	0.01				
ICC	0.05				
N_{id}	78				
Observations	147				

Table 11

2-year-olds, action-learning: Individual changes in baseline-corrected PTL for the action learning data

	baseline-corrected PTL at 2 years				
Predictors	*Estimates*	*std. Error*	*CI*	*Statistic*	*p*
(Intercept)	-0.018	0.060	-0.136 – 0.100	-0.296	0.767
Bayley percentage score	0.000	0.001	-0.001 – 0.002	0.291	0.771
Random Effects					
σ^2	0.08				
$\tau_{00\ id}$	0.00				
N_{id}	77				
Observations	142				

4.2.3. Violation of Expectation task

All 80 provided us at least three trials in the VOE task, resulting in a total of 617 trials (96.4% of all trials shown) after we removed the top and bottom 10% of overall differences between neighboring data points as well as invalid datapoints. An additional 0.22% of the pupil data were removed because 12-year-olds looked at stimuli for less than *M* - 3*SD* at a time, and another 5.62% of data points that were still missing after data interpolation across a moving window of four consecutive data points from the full data set, such that data from a total of 598 trials (95.44% of all trials shown) provided by 80 2-year-olds was entered the model (n_{action} = 301 trials, n_{word} = 297 trials).

The model was overall of good stability (i.e., close range of min and max estimates derived from model subsets when compared to the full model estimates), however, the range of minimum and maximum estimates for the random slope of *condition* within *participant ID* was wide (see Table B11), with the correlation parameter being -1. We hence fitted an additional model which was identical to the initial full model but lacking the correlation between random intercept of *participant ID* and the random slope for *match* and *condition* (see Table B12), which improved stability of the estimates for the random parameters *condition* and *match* (see Table B13). Since both the full and the null model were similar with regard to the model estimates, the following analyses were based on the pre-registered full model, see Table 12 for model estimates.

Table 12

2-year-olds: GLMM estimating effects of condition, temporal systematicity, match, and their interactions on pupil diameter changes

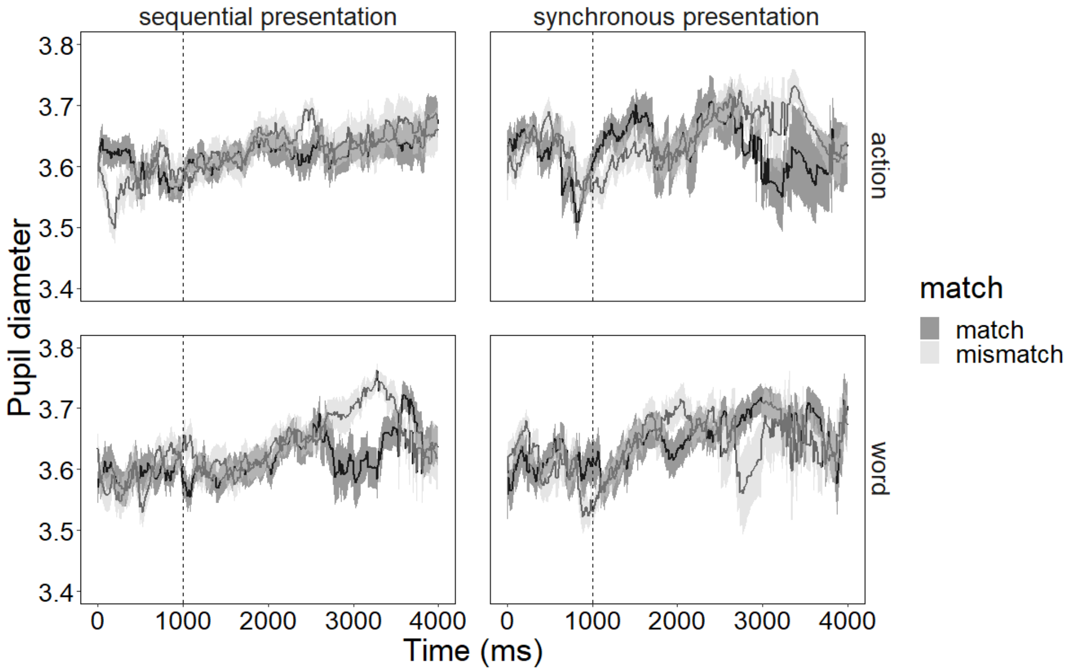
	baseline-corrected pupillary changes at 2 years				
Predictors	*Estimates*	*std. Error*	*CI*	*Statistic*	*p*
(Intercept)	0.679	0.027	0.626 – 0.732	25.072	**<0.001**
condition [word]	-0.075	0.031	-0.135 – -0.015	-2.451	0.014
match [mismatch]	-0.040	0.029	-0.097 – 0.017	-1.367	0.172
synchronicity [sync]	0.018	0.038	-0.056 – 0.093	0.482	0.630

condition [word] * match [mismatch]	-0.020	0.040	-0.099 – 0.059	-0.499	0.618
condition [word] * synchronicity [sync]	-0.042	0.043	-0.126 – 0.041	-0.997	0.319
match [mismatch] * synchronicity [sync]	-0.033	0.042	-0.114 – 0.049	-0.779	0.436
(condition [word] * match [mismatch]) * synchronicity [sync]	0.106	0.058	-0.007 – 0.219	1.831	0.067
Random Effects					
σ^2	0.03				
$\tau_{00\ id}$	0.00				
$\tau_{00\ id.1}$	0.01				
$\tau_{00\ item.x}$	0.00				
$\tau_{11\ id.condition.code}$	0.00				
$\tau_{11\ id.1.matchmismatch}$	0.00				
$\rho_{01\ id}$	-1.00				
$\rho_{01\ id.1}$	1.00				
N_{id}	80				
$N_{item.x}$	8				
Observations	598				

We inspected the data visually and concluded that all assumptions with regard to normal distributions and homogeneity of the residuals were met. The full-null model comparison revealed that the full model explained changes in 2-year-old's pupil dilations significantly better than the null-model ($\chi 2(7) = 22.607$, $p = .002$). We tested the contribution of individual fixed effects using drop1 (stats package version 4.0) which revealed that the 3-way-interaction term (including *condition*, *temporal systematicity* and *match*) was not significant ($p = 0.067$), see Figure 13. To further identify the extent to which individual predictors contributed to 2-year-old's response patterns, we simplified the model by taking out the interaction between *temporal systematicity* and *condition*, leaving *condition*, *match*, and their interaction as well as *temporal systematicity* and *match* and their interaction as additional fixed effect structures. Estimates of the simplified model revealed that 2-year-old's changes in pupil size were neither driven by differences in word- or action-cues for any of the match or mismatch trials ($p = 0.275$), nor by differences in the synchronous or sequential group for match or mismatch trials (p = 0.513), see Table B14.

Figure 13

2-year-olds: Changes in baseline-corrected pupil dilation for every combination of condition (word vs. action) and temporal systematicity (sequential and synchronous) in match and mismatch trials



Note. The dotted line indicates onset of label or action.

5. Discussion

Infants and young children learn about their world from a multisensory scene with words and actions often being presented together with objects. In the current study, we investigated the effect of temporal systematicity in word-action-object presentations on 1- and 2-year-olds' recognition of word-object and action-object mappings. Specifically, we tested infants' word-object and action-object learning when they had been presented with the information either in synchrony or in a sequential manner. We found evidence for learning of word-object and action-object mappings (as indexed by looking behavior) in the older age group, i.e., 2-year-olds, when the multisensory cues were presented in synchrony, but not when they were presented sequentially. We found no evidence for 1-year-olds' recognition of the previously trained mappings regardless of the temporal dynamics with which the triadic associations were presented during training. In what follows, we explain these findings in more detail.

Temporal systematicity

The data highlight the key role of multisensory synchrony on word-object and action-object learning in early development. The current study was motivated by conflicting findings in the literature with regards to children's learning performance in multisensory settings, i.e., settings where they are presented with both words and actions (e.g., Eiteljoerge et al., 2019; Puccini & Liszkowski, 2012). Eiteljoerge et al. (2019), for instance, suggested that the presentation of (non-aligned) cues from multiple modalities may be too demanding for infants and young children and hinder learning. We replicated this finding in the sequential condition across development (i.e., at 1- and 2-years of age). In particular, when children in the current study were presented with cues from multiple modalities in a sequential manner, neither age-group showed evidence of learning. In contrast, the older age-group showed evidence of learning when stimuli were presented in synchrony, highlighting that children benefit from temporally-aligned input during word-object and action-object learning.

Inconsistent looking time patterns across the two age groups are, furthermore, in line with research suggesting that the mechanisms underlying multisensory integration develop over the second year of life (Goldstein et al., 2020; Werker et al., 1998), when young children become more flexible in their learning strategies (Eiteljoerge et al., 2019a) and in interaction with their caregivers (Gogate et al., 2000a; Gogate & Bahrick, 2001). Unlike our expectations, 1-year-olds in this study did not show any benefits of temporally synchronous presentation of cues from multiple modalities nor did they show any evidence of having learned the word-object and action-object associations in the sequential condition. Our data are in line with results suggesting that multisensory processing in early development comes with additional costs, for example, that children do not show typically robust word-object association learning effects when words are accompanied by actions (Eiteljoerge et al., 2019a; Puccini & Liszkowski, 2012). Similarly, studies find that around 4 to 6 months of age, infants successfully match audio-visual content of rhesus monkey calls (Lewkowicz & Ghazanfar, 2006) as well as visual articulatory gestures and speech syllables (Pons et al., 2009), while older infants aged 8 to 11 months do not show evidence of such multisensory matching. The latter findings, in particular, have been taken to highlight processes of perceptual narrowing across early development, when mechanisms underlying the detection of frequent modal cues in the infant's immediate environment become more selective in their responsiveness to infrequent cues (Werker & Tees, 1984). During this sensitive developmental period, multisensory input may incur additional costs on mechanisms filtering modality-specific cues from word-action-object triads, leading to impoverished recognition of word-object and action-object mappings.

2-year-olds, in contrast, showed robust evidence of learning of both word-object and action-object associations when information from multiple modalities were presented in synchrony, i.e., in a temporally aligned manner, but not when they were presented sequentially, i.e., one after the other. Thus, in the synchronous condition, children looked significantly longer to the primed target object, i.e., the object that had been presented with the label or action prime at test. In the sequential group, while children looked - on average - longer at the target object preceded by a label-cue, they did not look significantly more at the target following presentation of the label-cue relative to prior presentation of this cue. The effect of temporal systematicity on learning is consistent with previous findings, for example, where infants look as they visually engage with an object can depend on whether an object is labeled in the moment that the object is visible (Althaus & Mareschal, 2014). Temporal alignment of multiple cues may help young children to focus on salient object features and, consequently, promote object encoding (Lany, Aguero, et al., 2022) which, in turn, facilitates object recognition and word-object learning. Additionally, temporally aligned cues may also help infants to draw symbolic connections between the object and the dynamic audio-visual information, potentially based on shape-centric object representations (Lany, et al., 2022). This result might also explain why Eiteljoerge et al., (2019) did not find any learning at two years given that they presented the information in a sequential manner. Young children may, thus, benefit from the temporal alignment of multisensory cues because temporal systematicity enables them to thoroughly and actively engage with the multisensory content in real-time as they effectively guide their attention between the physical object and the concurrent presentation of words and actions.

Benefits of temporal systematicity may also be mirrored at the level of neural processing. Indeed, the efficiency and robustness with which sensory stimuli are encoded is subject to the oscillatory activity of groups of neurons involved in a particular neural system (Peelle & Davis, 2012), i.e., the synchronization of firing patterns among neuronal groups. Such firing patterns in the sensory cortex align with the regular temporal structure of information in the input (Lakatos et al., 2008; Schroeder & Lakatos, 2009) and are systematically related to the speed and magnitude of response behavior (Lakatos et al., 2008). Temporal alignment of the input from different modalities would, therefore, create an optimal learning situation where processing efficiency is increased because neural oscillations across multiple systems line up with the temporal structure of the stimuli. Temporal systematicity in the input may therefore initiate highly-efficient processing opportunities via neural coupling (Peelle & Davis, 2012) as young children scan the object, hear the label and see the action performed on the object, such that robust associations are formed between the words, actions and objects. Processing of sequentially presented multisensory stimuli may, therefore, be less efficient in the sense that oscillatory waves don’t align completely to the temporal rhythm of the input, potentially leading to interruptions in processing of consecutively presented novel stimuli.

While children in the synchronous group were able to align to simultaneously presented information from multiple modalities, those in the sequential group had to sustain attention over two successive events, i.e., the action and the word for the object, in order to receive the entirety of information provided in the multisensory scene. Holding information about potential word-object associations available while simultaneously forming an action-object representation - and vice versa - may exceed working memory capacities to the extent that learning such associations is hindered. It may, therefore, be easier for the young brain to adapt to temporally-aligned sources of input with temporal alignment being more typical of caregiver-child interactions (Gogate et al., 2006; Matatyaho & Gogate, 2008) than temporal mis-alignment. Nevertheless, we note that currently, our study cannot rule out the possibility that sequential presentations of multisensory information may not necessarily hinder processing per se, as much as temporal alignment may boost processing.

Violation of expectation

Most studies investigating the effect of temporal dynamics in early developmental learning have used looking behavior as an index of learning which involves a voluntary bias of attention to features within the visual scene as young children encode, disambiguate, and retain novel perceptual and auditory links (Gogate, 2020; Gogate et al., 2000, 2001; Gogate & Bahrick, 1998; see introduction for a summary). In addition, we investigated differences in participants' pupillary response to matching and mismatching words and actions for objects as an additional index of learning, with the possibility that pupillary arousal may prove a more sensitive index of learning progress (Jackson & Sirois, 2009; Sirois & Jackson, 2012). However, we found no evidence of learning in either condition or age group in the violation of expectation task. On the one hand, this may speak to the robustness of the learned mappings with mismatches / violations being potentially more difficult to detect relative to previously-seen word-action-object mappings (c.f. Ballem & Plunkett, 2005, for a similar argument regarding performance in preferential looking and habituation tasks). On the other hand, this may also speak to the transience of the learned mappings, since the violation of expectation task was always presented after the looking time task. This was necessary to ensure that the mismatching trials did not impair learning progress and recognition in the looking behaviour task. A possibility, therefore, remains that the violation of expectation task would have indexed learning had it been presented immediately after the training task. Early in life, short attentional spans and working-memory limitations on the representation of novel information may impair the temporally-fragile associative links formed, leading to reduced recognition of word-object and action-object pairings when these cues are presented in isolation after a few minutes. Currently, our study cannot discriminate between these two interpretations of the results of the violation-of-expectation task.

Language proficiency

We did not find any association between individual differences in infants' and young children's developmental abilities including word production (infants and children) and word comprehension, fine-motor skills and gesture use (infants) and word-object or action-object recognition in synchronous or sequential settings. Parental estimates of children's receptive vocabulary have been shown to misalign with children's preferential looking behavior in word-learning tasks (Houston-Price et al., 2007), which might explain some of the inconsistencies between parental ratings of children's developmental milestones and recognition of newly learned associations in real-time. Moreover, reliable estimates of young children's associative learning success when assessing their language abilities via parental questionnaires such as the CDI can be challenging with regard to the early variability in developmental trajectories (Law & Roy, 2008) such that individual questionnaire scores may not reflect their potential in language development. In contrast, it may also be that parental reports did not accurately tap into artificially-generated word- and action-object settings because their reports subjectively summarize numerous situations in which they observe their child drawing connections among words, actions and objects across contexts and social interactions, while momentary looking time data indexes learning from a single testing event. The latter would suggest that individual effects in early language and action learning are difficult to capture using a between-participants design where children's looking behavior is usually tested once. Here, a repeated-measures design where children's target recognition patterns are tested multiple times might be more appropriate to disclose individual differences in early development on their word- and action-object learning success.

Word- and action-learning

Finally and contrary to our expectations, we found no evidence that infants' and young children's word-object and action-object association learning either differed across development or were differentially modulated by the temporal characteristics of stimulus presentation. At 1 year, infants did not show any evidence of recognition of the word-object or action-object associations at test, regardless of whether the stimuli were presented simultaneously or sequentially. This might be taken to suggest that early mechanisms underlying prelinguistic word-object and action-object associations respond similarly to temporal fluctuations in the input with regards to whether these cues are presented synchronously or sequentially. Equally, at 2 years of age, multisensory synchronicity facilitated learning of both – actions and labels for novel objects – while the sequential presentation of multisensory input hindered learning of both word- and action-object mappings. The influence of temporal alignment in the input, hence, highlights a potential similarity in the mechanisms underlying word-object and action-object encoding based on word- and action-cues. We take this to imply that characteristics of perceptual and cognitive processes involved in associating and recognizing objects with words and actions are likely to mature at a similar pace in early development, which further supports the idea that initial structures underlying such processes may be driven by overlapping or similar mechanisms. Presenting words and actions with a physical object assists in highlighting perceptual aspects and conceptual relations among object features and between object categories (e.g., Althaus & Mareschal, 2014; Goldstein et al., 2020), which, in turn, facilitates target recognition upon word and action primes to a similar extent. Word and action processing may overlap with regards to how they are associated with object features, and may merely be interpreted as means of communication between the agent and the observer.

Indeed, research comparing the emergence of words and actions in early development highlight the similarity in the timescales of action, gesture and lexical acquisition. For example, early gesture-word combinations precede the stringing of two words (Capirci et al., 2005), and infant's object exploration elicits caregiver's naming of the object, providing them with an opportunity to further conceptualize the physical object beyond action-related object examination (Suarez-Rivera et al., 2019). From an evolutionary point of view, actions have been suggested to derive from activation of the hand mirror system which later gives rise to the activation of related mirror systems involved in speech processing (Arbib, 2006; see also Goldin-Meadow et al., 1993), emphasizing how early concepts are first acquired by perceiving and actively manipulating an object which are then transferred into linguistic dimensions (e.g., Sloutsky & Napolitano, 2003) to extend object knowledge. If communicative behavior is indeed rooted in actions and gestures (Capirci et al., 2005; Caselli et al., 2012; Cattani et al., 2019), then word-object and action-object mappings should derive from shared mechanisms which operate under similar conditions, and here, we showed that they do with regard to the temporal manipulation of the multisensory word-action-object triads.

In conclusion, 1- and 2-year-old's learning about objects is embedded in a multisensory environment in which interaction partners often provide both language and action information about objects. In the current study, we showed that children at 2 years of age learn such information better when a word and an action for an object are presented in a temporally aligned manner. This was not the case for 1-year-olds who showed difficulties learning the novel word-action-object associations. The temporal dynamics in which these cues occur in the infant's immediate surroundings hence impact learning multisensory cues in early childhood.

Appendix A

Supplementary model information (1-year-olds)

This appendix provides estimates and output information for additional models that were either reduced versions of the models reported in the result section or updated versions of the registered models reported in this paper.

Table A 1

*Updated model estimating effects of condition * systematicity with adapted random effect structure*

	baseline-corrected PTL at 1 year				
Predictors	*Estimates*	*SE*	*CI*	*t*	*p*
(Intercept)	-0.02	0.04	-0.09 – 0.05	-0.48	0.634
condition [word]	-0.04	0.05	-0.15 – 0.06	-0.83	0.404
systematicity [sync]	0.00	0.05	-0.10 – 0.11	0.09	0.926
condition [word] * systematicity [sync]	0.04	0.07	-0.10 – 0.19	0.58	0.561
Random Effects					
σ^2	0.10				
$\tau_{00\ id}$	0.00				
$\tau_{00\ prime.used}$	0.00				
$\tau_{11\ id.condition.code}$	0.00				
$\tau_{11\ prime.used.systematicity.code}$	0.00				
ρ_{01}					
ρ_{01}					
N $_{id}$	78				
N $_{prime.used}$	4				
Observations	296				

Note. This table provides estimates for the model estimating the effects of condition and systematicity and their interactions as fixed effects and the random effect structures comprising the intercept and slope of condition (dummy coded & centered) within subject ID and intercept and slope of systematicity (dummy coded & centered) within primes shown at test (i.e., four levels, two actions and two labels). Log likelihood ratio tests of the updated model ($\chi2(11) = -80.209$) and a simpler version of the updated model (i.e., lacking intercept and slope correlations in the random effect structure and estimating all random effects to be 0, $\chi2(9) = -80.209$) revealed similar likelihood estimates such that the simpler model of the two was used for further analyses. Estimates for the updated model and the registered model reported here were identical ($\chi2(3) = 1.194$, $p = 0.754$).

Table A 2

Model stability estimates for the full GLMM testing condition, systematicity and their interactions

Predictor	*Estimate*	*min*	*max*
(Intercept)	-0.02	-0.03	-0.01
conditionword	-0.04	-0.07	-0.01
systematicitysync	0.01	-0.01	0.02
conditionword:systematicitysync	0.04	-0.04	0.12
Random effects			
id@(Intercept)	0.00	0.00	0.04
id@condition.code	0.00	0.00	0.12
id@(Intercept)@condition.code	NaN	-1.00	1.00
prime.used@(Intercept)	0.00	0.00	0.00
Residual	0.32	0.31	0.32

Note. Min and max values indicate stability bounds for model estimates, derived by dropping data from one infant at a time from the full data set while comparing the model estimates from the resulting subsets with the estimates for the full data set.

Time-course analysis

We examined the data across the course of the trial by including the factor *time* and it's first, second and third polynomial. Apart from some expected deviations of homogeneity assumptions for residuals plotted against fitted values, visual data inspection proved that assumptions for normally distributed and homogeneous residuals were met. The full-null model comparison showed that the full model comprising all predictors (for estimates see Table A3) explained change in infant's baseline-corrected PTL significantly better than the null model lacking condition and temporal systematicity ($\chi2(12) = 42.402$, $p < .001$). Testing the effect of individual fixed effects suggested that condition interacting with temporal systematicity and time (i.e., first, second and third polynomial) significantly changed infants' looking time response over the course of the test trials ($p < .001$).

Table A 3

1-year-olds: GLMM estimating effects of condition, systematicity, time and their interactions

	baseline-corrected PTL at 1 year			
Predictors	*Estimate*	*Std.Error*	*t-value*	*p-value*
(Intercept)	-0.02	0.03	-0.68	0.50
conditionword	-0.03	0.02	-1.33	0.18
systematicitysync	0.01	0.04	0.16	0.87
ot1	-3.67	2.95	-1.24	0.21
ot2	2.16	2.47	0.88	0.38
ot3	2.78	1.99	1.40	0.16
conditionword:systematicitysync	0.00	0.02	0.27	0.79
conditionword:ot1	6.11	2.90	2.11	0.04
conditionword:ot2	-1.45	1.68	-0.86	0.39
conditionword:ot3	-3.80	1.46	-2.61	**<0.001**
systematicitysync:ot1	5.94	3.34	1.78	0.08
systematicitysync:ot2	-2.89	3.42	-0.84	0.40
systematicitysync:ot3	-0.42	2.83	-0.15	0.88
conditionword:systematicitysync:ot1	-5.94	2.02	-2.95	**<.001**
conditionword:systematicitysync:ot2	3.57	2.07	1.73	0.08
conditionword:systematicitysync:ot3	2.90	2.00	1.45	0.15

Table A 4

Model stability estimates for the full GLMM testing condition, systematicity, time and their interactions

Predictor	*Estimate*	*min*	*max*
(Intercept)	-0.02	-0.03	-0.00
conditionword	-0.03	-0.04	-0.01
systematicitysync	0.01	-0.02	0.03
ot1	-3.67	-4.52	-2.07
ot2	2.16	0.81	3.99
ot3	2.78	-0.25	5.78
conditionword:systematicitysync	0.00	-0.05	0.05
conditionword:ot1	6.11	4.20	8.72
conditionword:ot2	-1.45	-3.57	1.05
conditionword:ot3	-3.80	-6.60	1.00
systematicitysync:ot1	5.94	-3.73	13.26
systematicitysync:ot2	-2.89	-7.08	-1.13
systematicitysync:ot3	-0.42	-5.33	2.55
conditionword:systematicitysync:ot1	-5.94	-14.95	4.20
conditionword:systematicitysync:ot2	3.57	-3.73	10.58
conditionword:systematicitysync:ot3	2.90	-0.00	7.11
Random effects			
id@(Intercept)	0.15	0.14	0.19
id@ot1	13.50	12.92	17.09
id@ot2	13.83	4.60	18.57
id@ot3	10.97	10.15	13.73
id@(Intercept)@ot1	0.11	-0.00	0.30
id@(Intercept)@ot2	0.05	-0.33	0.18
id@(Intercept)@ot3	-0.19	-0.35	0.10
id@ot1@ot2	0.03	-0.96	0.16

id@ot1@ot3	-0.23	-0.40	-0.07
id@ot2@ot3	0.33	0.23	0.46
prime.used@(Intercept)	0.02	0.00	0.02
prime.used@ot1	2.56	0.10	3.26
prime.used@ot2	0.96	0.22	1.73
prime.used@ot3	0.55	0.00	0.92
prime.used@(Intercept)@ot1	0.83	-1.00	1.00
prime.used@(Intercept)@ot2	0.67	-1.00	1.00
prime.used@(Intercept)@ot3	0.66	-1.00	1.00
prime.used@ot1@ot2	0.97	0.81	1.00
prime.used@ot1@ot3	0.12	-1.00	1.00
prime.used@ot2@ot3	-0.12	-1.00	1.00
Residual	0.45	0.42	0.47

Note. Model stability shown as min and max values for model estimates which were derived by dropping data from individuals (one at a time) from the full data set while comparing these model estimates derived from the subsets with the estimates for the full data set.

To further identify the contribution of individual fixed effects, we reduced the full model to test fixed effect combinations for lower order terms whereby the random effect structure remained unchanged. We first reduced the 3-way-interaction term including *condition, temporal systematicity* and *time* (i.e., all three polynomials) to *condition*, *temporal systematicity* and the *first and second polynomial* (i.e., dropping the third polynomial from the interaction term), while keeping a squared combination of *condition*, *temporal systematicity* and the *third polynomial* in the model. The *third polynomial* interacting with *condition* (p = 0.272) and temporal systematicity (p = 0.752) remained non-significant and was hence dropped from the model (see Table A5), which left us with a reduced model including the shortened 3-way-interaction (including condition, temporal systematicity and the first and second polynomial) and the single third polynomial as an additional fixed effect.

Table A 5

Reduced time-course model version a

Predictors	*Estimate*	*SE*	*df*	*t.value*	*p.value*
(Intercept)	-0.02	0.02	82.88	-0.71	0.48
conditionword	-0.03	0.01	11246.91	-2.23	0.03
systematicitysync	0.00	0.04	84.39	0.15	0.88
ot[, 1:2]1	-3.56	3.34	149.29	-1.07	0.29
ot[, 1:2]2	2.16	2.57	48.51	0.84	0.40
ot[, 3]	2.27	2.02	21.40	1.12	0.28
conditionword:systematicitysync	0.00	0.02	11250.98	0.16	0.87
conditionword:ot[, 1:2]1	6.16	3.62	94.78	1.70	0.09
conditionword:ot[, 1:2]2	-1.21	1.92	7.45	-0.63	0.55
systematicitysync:ot[, 1:2]1	6.09	3.38	87.59	1.80	0.07
systematicitysync:ot[, 1:2]2	-2.62	3.45	82.80	-0.76	0.45
conditionword:ot[, 3]	-2.48	1.30	1.21	-1.91	0.27
systematicitysync:ot[, 3]	0.86	2.71	70.01	0.32	0.75
conditionword:systematicitysync:ot[, 1:2]1	-6.43	2.00	10976.81	-3.22	**<0.001**
conditionword:systematicitysync:ot[, 1:2]2	2.87	2.04	10781.60	1.41	0.16

Note. The reduced model lacks the third polynomial in the critical 3-way-interaction; i.e. the model estimates the effects of condition, systematicity, and first and second polynomial as well as their interactions and additionally treats the squared sum of condition, systematicity and the third polynomial as additional fixed effects.

The main effect of the third polynomial was not informative with regard to infants' change in baseline-corrected PTL (p = 0.357) and was hence dropped from the model (see Table A6).

Table A 6

Reduced time-course model version b

Predictors	*Estimate*	*SE*	*df*	*t.value*	*p.value*
(Intercept)	-0.03	0.03	16.68	-1.03	0.32
conditionword	-0.00	0.02	3.02	-0.20	0.85
systematicitysync	0.01	0.04	83.54	0.21	0.83
ot[, 1:2]1	-4.00	3.37	6.27	-1.19	0.28
ot[, 1:2]2	2.45	2.52	36.18	0.97	0.34
ot[, 3]	1.43	1.52	24.02	0.94	0.36
conditionword:systematicitysync	0.00	0.02	11281.10	0.17	0.86
conditionword:ot[, 1:2]1	6.76	3.68	2.31	1.83	0.19
conditionword:ot[, 1:2]2	-1.48	1.88	4.33	-0.78	0.47
systematicitysync:ot[, 1:2]1	6.12	3.34	84.26	1.83	0.07
systematicitysync:ot[, 1:2]2	-3.20	3.30	90.49	-0.97	0.33
conditionword:systematicitysync:ot[, 1:2]1	-6.25	2.00	10980.80	-3.13	**<0.001**
conditionword:systematicitysync:ot[, 1:2]2	3.15	2.03	10730.44	1.55	0.12

Note. The reduced model lacks the third polynomial in the critical 3-way-interaction; i.e. the model estimates the effects of condition, systematicity, and first and second polynomial as well as their interactions and additionally treats the third polynomial as additional fixed effect.

The further reduced model comprising condition, temporal systematicity and the first and second polynomial and their interactions was significant for the first polynomial interaction term (p = 0.002), but not the second polynomial interaction term (p = 0.12), which is why the latter was dropped from the model (see Table A7).

Table A 7

Reduced time-course model version c

Predictors	*Estimate*	*SE*	*df*	*t.value*	*p.value*
(Intercept)	-0.03	0.03	18.21	-1.10	0.29
conditionword	-0.00	0.02	3.03	-0.18	0.87
systematicitysync	0.01	0.04	83.49	0.20	0.84
ot[, 1:2]1	-3.85	3.37	6.12	-1.14	0.30
ot[, 1:2]2	2.08	2.50	32.01	0.83	0.41
conditionword:systematicitysync	0.00	0.02	11281.31	0.17	0.86
conditionword:ot[, 1:2]1	6.79	3.69	2.30	1.84	0.19
conditionword:ot[, 1:2]2	-1.50	1.89	4.29	-0.79	0.47
systematicitysync:ot[, 1:2]1	6.05	3.34	84.25	1.81	0.07
systematicitysync:ot[, 1:2]2	-3.25	3.30	90.45	-0.99	0.33
conditionword:systematicitysync:ot[, 1:2]1	-6.25	2.00	10991.27	-3.13	**<0.001**
conditionword:systematicitysync:ot[, 1:2]2	3.16	2.03	10746.61	1.56	0.12

Note. The reduced model lacks the third polynomial in the critical 3-way-interaction; i.e. the model estimates the effects of condition, systematicity, and first and second polynomial and their interactions.

To further simplify the model, the 3-way-interaction term now included condition, temporal systematicity and the first polynomial and their interactions (i.e., lacking the second and third polynomial), while adding the squared sum of condition, temporal systematicity and the second polynomial as fixed effects (see Table A8).

Table A 8

Reduced time-course model version d

Predictors	*Estimate*	*SE*	*df*	*t.value*	*p.value*
(Intercept)	-0.03	0.03	18.00	-1.09	0.29
conditionword	-0.00	0.02	3.00	-0.15	0.89
systematicitysync	0.01	0.04	83.51	0.21	0.84
ot[, 1]	-3.75	3.37	6.14	-1.11	0.31
ot[, 2]	1.51	2.46	29.66	0.61	0.54
conditionword:systematicitysync	0.00	0.02	11274.14	0.04	0.97
conditionword:ot[, 1]	6.83	3.69	2.30	1.85	0.19
systematicitysync:ot[, 1]	6.13	3.34	84.38	1.83	0.07
conditionword:ot[, 2]	-0.16	1.65	2.53	-0.10	0.93
systematicitysync:ot[, 2]	-1.80	3.15	76.99	-0.57	0.57
conditionword:systematicitysync:ot[, 1]	-6.75	1.97	11192.89	-3.43	**<0.001**

Note. The reduced model lacks the second and third polynomial in the critical 3-way-interaction; i.e. the model estimates the effects of condition, systematicity and the first polynomial and their interactions as well as the squared sum of condition, systematicity and the second polynomial as additional fixed effect term.

Interactions between the second polynomial and condition (p = 0.929) or temporal systematicity (p = 0.569) did not explain considerable changes in infants' baseline-corrected PTL, which hence left us with an even simpler model including condition, temporal systematicity and the first polynomial and their interactions and the second polynomial as additional fixed effect (see Table A9), and this additional fixed effect term also remained non-significant to the response (p = 0.758).

Table A 9

Reduced time-course model version e

Predictors	*Estimate*	*SE*	*df*	*t.value*	*p.value*
(Intercept)	-0.02	0.03	34.61	-0.86	0.40
conditionword	-0.02	0.02	5.52	-0.98	0.37
systematicitysync	0.01	0.04	84.24	0.27	0.79
ot[, 1]	-4.50	3.06	12.60	-1.47	0.17
ot[, 2]	0.52	1.66	46.53	0.31	0.76
conditionword:systematicitysync	0.00	0.02	11259.08	0.05	0.96
conditionword:ot[, 1]	7.74	2.64	4.23	2.93	**0.04**
systematicitysync:ot[, 1]	6.35	3.32	86.41	1.91	0.06
conditionword:systematicitysync:ot[, 1]	-6.70	1.97	11021.14	-3.40	**<0.001**

Note. The reduced model lacks the second and third polynomial in the critical 3-way-interaction, i.e. the model estimates the effects of condition, systematicity and the first polynomial and their interactions as well as the second polynomial as additional fixed.

The final reduced version of the full model comprised condition, temporal systematicity and the first polynomial and their interactions in the fixed effects structure (see Table A10) and predicted infants' linear looking time patterns over the course of the test trials (t = -3.448, SE = 1.97, p = .001).

Table A 10

Reduced time-course model version f

	baseline-corrected PTL at 1 year				
Predictors	*Estimate*	*SE*	*df*	*t.value*	*p.value*
(Intercept)	-0.02	0.02	82.98	-0.69	0.49
conditionword	-0.03	0.01	1618.78	-2.19	**0.03**
systematicitysync	0.01	0.04	84.14	0.25	0.80
ot[, 1]	-3.88	2.98	10.90	-1.30	0.22
conditionword:systematicitysync	0.00	0.02	11206.55	0.06	0.95
conditionword:ot[, 1]	6.71	2.59	3.77	2.59	0.06
systematicitysync:ot[, 1]	6.38	3.32	86.43	1.92	0.06
conditionword:systematicitysync:ot[, 1]	-6.79	1.97	10909.62	-3.45	**<0.001**

Note. The final reduced model lacks the second and third polynomial in the critical 3-way-interaction; i.e. the model estimates the effects of condition, systematicity and the first polynomial and their interactions.

Table A 11

Model stability for ELFRA score estimates in the word learning data

Predictor	*Estimate*	*min*	*max*
(Intercept)	-0.08	-0.13	-0.04
ELFRAcompr	0.00	0.00	0.00
ELFRAprod	0.01	0.01	0.01
ELFRAgestures	-0.00	-0.00	0.00
ELFRAfmot	-0.01	-0.01	0.00
id@(Intercept)	0.00	0.00	0.00
Residual	0.32	0.30	0.32

Table A 12

Model stability for ELFRA score estimates in the action learning data

Predictor	*Estimate*	*min*	*max*
(Intercept)	-0.08	-0.13	-0.04
ELFRAcompr	0.00	0.00	0.00
ELFRAprod	0.01	0.01	0.01
ELFRAgestures	-0.00	-0.00	0.00
ELFRAfmot	-0.01	-0.01	0.00
id@(Intercept)	0.00	0.00	0.00
Residual	0.32	0.30	0.32

Table A 13

Model stability estimates derived from testing condition, match, systematicity and their interactions

	baseline-corrected pupil dilation at 1 year		
Predictor	*Estimate*	*min*	*max*
(Intercept)	0.56	0.53	0.58
conditionword	-0.01	-0.04	0.01
matchmismatch	0.02	-0.00	0.05
systematicitysync	0.01	-0.01	0.02
conditionword:matchmismatch	-0.00	-0.03	0.02
conditionword:systematicitysync	0.01	-0.00	0.02
matchmismatch:systematicitysync	0.01	-0.04	0.03
conditionword:matchmismatch:systematicitysync	-0.03	-0.06	0.02
Random effects			
id@(Intercept)	0.13	0.10	0.13
id@condition.code	0.05	0.04	0.06
id@(Intercept)@condition.code	-1.00	-1.00	-1.00
id@matchmismatch	0.06	0.05	0.08

id@(Intercept)@matchmismatch	-1.00	-1.00	-1.00
item.x@(Intercept)	0.03	0.00	0.03
Residual	0.17	0.16	0.17

Appendix B

Supplementary model information (2-year-olds)

This appendix provides estimates and output information for additional models that were either reduced versions of the models reported in the result section or updated versions of the registered models reported in this paper.

Table B 1

*Updated model estimating effects of condition * systematicity with adapted random effect structure*

Predictors	*Estimate*	*Std.Error*	*df*	*t.value*	*p.value*
(Intercept)	-0.09	0.04	293	-2.49	**0.01**
conditionword	0.10	0.05	293	2.05	**0.04**
systematicitysync	0.18	0.05	293	3.45	**<0.001**
conditionword:systematicitysync	-0.12	0.07	293	-1.62	0.11

Note. This table provides estimates for the model estimating the effects of condition and systematicity and their interactions as fixed effects and the random effect structures comprising the intercept and slope of condition (dummy coded & centered) within subject ID and intercept and slope of systematicity (dummy coded & centered) within primes shown at test (i.e., four levels, two actions and two labels). Log likelihood ratio tests of the updated model ($\chi 2(11) = -68.97$) and a simpler version of the updated model (i.e., lacking intercept and slope correlations in the random effect structure and estimating all random effects to be 0, $\chi 2(9) = -70.49$) revealed similar likelihood estimates such that the simpler model of the two was used for further analyses. Estimates for the updated model and the registered model reported here were similar despite random effect structure adjustment ($p = 0.106$).

Table B 2

Model stability estimates for the full GLMM testing condition, systematicity and their interactions

Predictor	*Estimate*	*min*	*max*
(Intercept)	-0.09	-0.11	-0.07
conditionword	0.10	0.08	0.13
systematicitysync	0.18	0.13	0.22
conditionword:systematicitysync	-0.12	-0.17	-0.06
id@(Intercept)	0.06	0.03	0.12
id@condition.code	0.14	0.10	0.24
id@(Intercept)@condition.code	1.00	1.00	1.00
prime.used@(Intercept)	0.00	0.00	0.00
Residual	0.29	0.27	0.30

Note. Min and max values indicate stability bounds for model estimates, derived by dropping data for one child at a time from the full data set while comparing the model estimates from the resulting subsets with the estimates for the full data set.

Table B 3

2-year-olds: GLMM testing the main effects of condition and systematicity

Predictors	*Estimate*	*Std.Error*	*t.value*	*p.value*
(Intercept)	-0.07	0.03	-2.15	**0.03**
conditionword	0.05	0.04	1.21	0.23
systematicitysync	0.13	0.04	3.48	**<0.001**

Time-course analysis

We examined the effect of *condition* and *temporal systematicity* interacting with the factor *time* and it’s first, second and third polynomial within the model to examine children’s looking time patterns over the course of the test trials. As with the 1-year-olds, there was considerable overlap between the results of the model including and excluding time. Homogeneity assumptions for residuals plotted against fitted values were deviated but expected, and visual inspection of data showed that the data were normally distributed and residuals were assumed to be homogeneous. The full model (containing the 3-way-interaction, for estimates see Table B4) explained changes in children's looking patterns over the course of the trial significantly better than the null model (lacking condition and temporal systematicity; $\chi 2(12) = 59.869$, $p < .0001$), and the full model had good stability (see Table B5).

Table B 4

2-year-olds: GLMM estimating effects of condition, systematicity, time and their interactions

Predictors	*Estimate*	*Std.Error*	*t.value*	*p.value*
(Intercept)	-0.08	0.03	-2.96	**<0.001**
conditionword	0.04	0.01	3.34	**<0.001**
systematicitysync	0.17	0.04	4.76	**<0.001**
ot1	-2.72	2.49	-1.09	0.27
ot2	-1.29	2.36	-0.55	0.58
ot3	2.03	2.20	0.92	0.36
conditionword:systematicitysync	-0.06	0.02	-3.19	**<0.001**
conditionword:ot1	1.20	1.46	0.82	0.41
conditionword:ot2	-1.81	1.75	-1.03	0.30
conditionword:ot3	-4.59	2.07	-2.22	**0.03**
systematicitysync:ot1	4.77	3.51	1.36	0.17
systematicitysync:ot2	4.07	3.21	1.27	0.20
systematicitysync:ot3	-1.47	2.76	-0.53	0.59
conditionword:systematicitysync:ot1	0.53	2.05	0.26	0.80
conditionword:systematicitysync:ot2	-3.17	2.10	-1.50	0.13
conditionword:systematicitysync:ot3	8.52	2.08	4.09	**<0.001**

Table B 5

Model stability estimates for the full GLMM testing condition, systematicity, time and their interactions

Predictor	*Estimate*	*min*	*max*
(Intercept)	-0.08	-0.09	-0.06
conditionword	0.04	0.02	0.07
systematicitysync	0.17	0.16	0.19
ot1	-2.72	-3.91	-1.62
ot2	-1.29	-2.67	-0.05
ot3	2.03	-1.32	6.49
conditionword:systematicitysync	-0.06	-0.11	-0.01
conditionword:ot1	1.20	-2.39	3.21
conditionword:ot2	-1.81	-8.75	3.42
conditionword:ot3	-4.59	-10.37	-0.49
systematicitysync:ot1	4.77	3.44	6.17
systematicitysync:ot2	4.07	0.89	5.46
systematicitysync:ot3	-1.47	-7.15	3.17
conditionword:systematicitysync:ot1	0.53	-2.52	5.18
conditionword:systematicitysync:ot2	-3.17	-10.54	4.00
conditionword:systematicitysync:ot3	8.52	3.32	15.51
Random effects			
id@(Intercept)	0.15	0.14	0.19
id@ot1	14.23	9.76	18.08
id@ot2	12.62	8.38	16.02
id@ot3	10.37	9.80	12.97
id@(Intercept)@ot1	0.43	0.31	1.00
id@(Intercept)@ot2	-0.01	-0.30	0.14
id@(Intercept)@ot3	-0.30	-0.45	-0.16
id@ot1@ot2	0.08	-0.01	0.61

id@ot1@ot3	-0.49	-0.54	-0.32
id@ot2@ot3	0.12	-0.07	0.19
prime.used@(Intercept)	0.00	0.00	0.01
prime.used@ot1	0.29	0.01	3.10
prime.used@ot2	1.00	0.09	4.69
prime.used@ot3	1.49	0.74	6.79
prime.used@(Intercept)@ot1	NaN	-1.00	1.00
prime.used@(Intercept)@ot2	NaN	-1.00	0.31
prime.used@(Intercept)@ot3	NaN	-1.00	1.00
prime.used@ot1@ot2	0.97	-0.86	1.00
prime.used@ot1@ot3	-0.10	-1.00	1.00
prime.used@ot2@ot3	-0.27	-1.00	1.00
Residual	0.47	0.44	0.48

Because condition interacting with temporal systematicity and time (i.e., first, second and third polynomial) significantly impacted the response pattern ($p < .0001$), we reduced the full model to test the 3-way-interaction with lower order terms (random effect structure remained unchanged). We first reduced the initial 3-way-interaction term to condition, temporal systematicity and the first and second polynomial and their interactions (i.e., dropping the third polynomial from the interaction term) and added a squared combination of condition, temporal systematicity and the third polynomial to the model. The third polynomial interacting with condition ($p = 0.814$) and temporal systematicity ($p = 0.32$) did not correspond to changes in the response, so these combinations were dropped from the model (see Table B6).

Table B 6

Reduced time-course model version a

Predictors	*Estimate*	*SE*	*df*	*t.value*	*p.value*
(Intercept)	-0.08	0.03	82.75	-2.93	**<0.001**
conditionword	0.04	0.01	11114.33	3.39	**<0.001**
systematicitysync	0.17	0.04	82.95	4.75	**<0.001**
ot[, 1:2]1	-2.72	2.54	76.25	-1.07	0.29
ot[, 1:2]2	-1.29	2.47	26.24	-0.52	0.60
ot[, 3]	0.24	2.38	9.39	0.10	0.92
conditionword:systematicitysync	-0.06	0.02	11080.80	-3.41	**<0.001**
conditionword:ot[, 1:2]1	1.27	1.49	19.69	0.85	0.41
conditionword:ot[, 1:2]2	-1.56	1.95	3.38	-0.80	0.48
systematicitysync:ot[, 1:2]1	5.17	3.58	84.06	1.45	0.15
systematicitysync:ot[, 1:2]2	4.59	3.25	86.33	1.41	0.16
conditionword:ot[, 3]	-0.60	2.27	2.07	-0.27	0.81
systematicitysync:ot[, 3]	2.61	2.60	74.83	1.00	0.32
conditionword:systematicitysync:ot[, 1:2]1	-0.33	2.04	9516.57	-0.16	0.87
conditionword:systematicitysync:ot[, 1:2]2	-4.50	2.08	8717.22	-2.16	**0.03**

Note. The reduced model lacks the third polynomial in the critical 3-way-interaction; i.e. the model estimates the effects of condition, systematicity, and first and second polynomial as well as their interactions and additionally treats the squared sum of condition, systematicity and the third polynomial as additional fixed effects.

We hence further reduced the model by keeping the shortened 3-way-interaction (including condition, temporal systematicity and the first and second polynomial) and added the single third polynomial as an additional fixed effect. The main effect of the third polynomial was not informative with regard to infants' change in baseline-corrected PTL (p = 0.43) and was hence dropped from the model (see Table B7).

Table B 7

Reduced time-course model version b

Predictors	*Estimate*	*SE*	*df*	*t.value*	*p.value*
(Intercept)	-0.08	0.03	84.09	-3.13	**<0.001**
conditionword	0.04	0.01	11011.00	3.39	**<0.001**
systematicitysync	0.18	0.04	81.47	5.10	**<0.001**
ot[, 1:2]1	-3.40	2.46	79.62	-1.38	0.17
ot[, 1:2]2	-1.04	2.45	27.01	-0.43	0.67
ot[, 3]	1.23	1.53	17.66	0.81	0.43
conditionword:systematicitysync	-0.06	0.02	10952.07	-3.40	**<0.001**
conditionword:ot[, 1:2]1	1.29	1.49	22.43	0.86	0.40
conditionword:ot[, 1:2]2	-1.63	1.91	3.43	-0.85	0.45
systematicitysync:ot[, 1:2]1	6.50	3.32	82.68	1.96	**0.05**
systematicitysync:ot[, 1:2]2	4.16	3.22	88.58	1.29	0.20
conditionword:systematicitysync:ot[, 1:2]1	-0.33	2.04	9114.53	-0.16	0.87
conditionword:systematicitysync:ot[, 1:2]2	-4.52	2.08	8550.48	-2.17	**0.03**

Note. The reduced model lacks the third polynomial in the critical 3-way-interaction; i.e. the model estimates the effects of condition, systematicity, and first and second polynomial as well as their interactions and treats the third polynomial as an additional fixed effect.

Solely keeping condition, temporal systematicity, as well as the first and second polynomial and their interactions as fixed effects significantly predicted response patterns for the second polynomial interaction term (p = 0.03), but not the first polynomial interaction term (p = 0.871, see Table B8). Since the second polynomial cannot be treated without its lower order terms (i.e., first polynomials) this final model cannot be reduced further with regard to the time components in the fixed effect structure and will serve to visualize the interacting contribution of condition, temporal systematicity and the first and second polynomial on children's target looking responses, see Figure B1. Confidence intervals for the final reduced model estimates were obtained using a parametric bootstrap (function bootMer of the package lme4; *N* = 1.000 bootstraps).

Table B 8

Reduced time-course model version c

Predictors	*Estimate*	*SE*	*df*	*t.value*	*p.value*
(Intercept)	-0.08	0.03	82.30	-3.04	**<0.001**
conditionword	0.04	0.01	11114.15	3.39	**<0.001**
systematicitysync	0.18	0.04	81.41	5.10	**<0.001**
ot[, 1:2]1	-2.95	2.39	66.40	-1.23	0.22
ot[, 1:2]2	-1.09	2.45	27.12	-0.45	0.66
conditionword:systematicitysync	-0.06	0.02	10861.96	-3.41	**<0.001**
conditionword:ot[, 1:2]1	1.30	1.49	21.79	0.87	0.39
conditionword:ot[, 1:2]2	-1.65	1.91	3.40	-0.86	0.44
systematicitysync:ot[, 1:2]1	6.50	3.32	82.68	1.96	**0.05**
systematicitysync:ot[, 1:2]2	4.15	3.22	88.43	1.29	0.20
conditionword:systematicitysync:ot[, 1:2]1	-0.33	2.04	8941.00	-0.16	0.87
conditionword:systematicitysync:ot[, 1:2]2	-4.51	2.08	8513.91	-2.17	**0.03**

Note. The reduced model lacks the third polynomial in the critical 3-way-interaction; i.e. the model estimates the effects of condition, systematicity, and first and second polynomial and their interactions.

Table B 9

Model stability for BabyLex percentage estimates in the word learning data

Predictor	*Estimate*	*min*	*max*
(Intercept)	0.03	0.01	0.06
BabyLexpercentage	0.00	-0.00	0.00
id@(Intercept)	0.08	0.02	0.10
Residual	0.33	0.32	0.33

Table B 10

Model stability for BabyLex percentage estimates in the action learning data

Predictor	*Estimate*	*min*	*max*
(Intercept)	-0.02	-0.04	0.02
Bperc	0.00	-0.00	0.00
id@(Intercept)	0.00	0.00	0.00
Residual	0.29	0.28	0.29

Table B 11

Model stability estimates for condition, systematicity, match and their interactions

Predictor	*Estimate*	*min*	*max*
(Intercept)	0.68	0.67	0.69
conditionword	-0.08	-0.09	-0.07
matchmismatch	-0.04	-0.07	-0.01
systematicitysync	0.02	0.01	0.04
conditionword:matchmismatch	-0.02	-0.05	0.01
conditionword:systematicitysync	-0.04	-0.06	-0.03
matchmismatch:systematicitysync	-0.03	-0.05	-0.01
conditionword:matchmismatch:systematicitysync	0.11	0.08	0.12
Random effects			
id@(Intercept)	0.01	0.00	0.11
id@condition.code	0.05	0.02	0.08
id@(Intercept)@condition.code	-1.00	-1.00	1.00
id@matchmismatch	0.04	0.03	0.11
id@(Intercept)@matchmismatch	1.00	0.89	1.00
item.x@(Intercept)	0.01	0.00	0.02
Residual	0.18	0.16	0.18

Note. Min and max values indicate stability bounds for model estimates, derived by dropping data for one child at a time from the full data set while comparing the model estimates from the resulting subsets with the estimates for the full data set.

Table B 12

GLMM estimating effects of condition, systematicity, match and their interactions with adapted random effect structure

	baseline-corrected pupil dilation at two-years				
Predictors	*Estimates*	*std. Error*	*CI*	*Statistic*	*p*
(Intercept)	0.68	0.03	0.63 – 0.73	25.22	**<0.001**
condition [word]	-0.08	0.03	-0.14 – -0.02	-2.45	**0.014**
match [mismatch]	-0.04	0.03	-0.10 – 0.02	-1.37	0.172
systematicity [sync]	0.02	0.04	-0.06 – 0.09	0.48	0.629
condition [word] * match [mismatch]	-0.02	0.04	-0.10 – 0.06	-0.50	0.618
condition [word] * systematicity [sync]	-0.04	0.04	-0.13 – 0.04	-1.00	0.319
match [mismatch] * systematicity [sync]	-0.03	0.04	-0.11 – 0.05	-0.78	0.436
(condition [word] * match [mismatch]) * systematicity [sync]	0.11	0.06	-0.01 – 0.22	1.83	0.067
N_{id}	80				
$N_{item.x}$	8				
Observations	598				

Note. The model was identical to the full model (see Table 9) while lacking correlations of random slope of condition and match within subject id.

Table B 13

Model stability estimates for condition, systematicity, match and their interactions w/o correlations of random slope of condition & match and random intercept of subject id

Predictor	*Estimate*	*min*	*max*
(Intercept)	0.68	0.67	0.69
conditionword	-0.08	-0.09	-0.07
matchmismatch	-0.04	-0.07	-0.01
systematicitysync	0.02	0.01	0.04
conditionword:matchmismatch	-0.02	-0.04	0.01
conditionword:systematicitysync	-0.04	-0.06	-0.03
matchmismatch:systematicitysync	-0.03	-0.05	-0.01
conditionword:matchmismatch:systematicitysync	0.11	0.08	0.12
id@(Intercept)	0.00	0.00	0.12
id@matchmatch	0.11	0.00	0.12
id@matchmismatch	0.15	0.04	0.15
id@matchmatch@matchmismatch	1.00	0.93	1.00
id@condition.code	0.05	0.00	0.08
item.x@(Intercept)	0.01	0.00	0.02
Residual	0.18	0.16	0.18

Note. Model stability for the random slope of condition was improved by taking out the correlation of random slope of condition and match within subject id.

Table B 14

Reduced pupil dilation model version a

	baseline-corrected pupil dilation at two-years				
Predictors	*Estimates*	*SE*	*CI*	*t*	*p*
(Intercept)	0.69	0.03	0.64 – 0.74	27.52	**<0.001**
condition [word]	-0.10	0.02	-0.14 – -0.05	-4.27	**<0.001**
systematicity [sync]	-0.00	0.03	-0.06 – 0.06	-0.08	0.933
match [mismatch]	-0.07	0.03	-0.12 – -0.02	-2.55	**0.011**
condition [word] * match [mismatch]	0.03	0.03	-0.03 – 0.09	1.09	0.275
systematicity [sync] * match [mismatch]	0.02	0.03	-0.04 – 0.08	0.65	0.513
Random Effects					
σ^2	0.03				
$\tau_{00\ id}$	0.00				
$\tau_{00\ id.1}$	0.01				
$\tau_{00\ item.x}$	0.00				
$\tau_{11\ id.condition.code}$	0.00				
$\tau_{11\ id.1.matchmismatch}$	0.00				
$\rho_{01\ id}$	-1.00				
$\rho_{01\ id.1}$	1.00				
N_{id}	80				
$N_{item.x}$	8				
Observations	598				

Note. The reduced model lacks the interaction component between condition and systematicity, leaving condition and match trials and their interaction as well as systematicity and match trials and their interaction as fixed effects structure.

Chapter 3

Can words do more than arbitrary actions in early category formation tasks?

Ricarda Bothe[1,3], Léonie Trouillet[2], Birgit Elsner[2] & Nivedita Mani[1,3]
[1]Georg-August University Göttingen, [2]University of Potsdam, [3]Leibniz Science Campus

Author statement

Data presented in the following article will serve to address two distinct research questions. First, we will directly compare the data collected online and via eye tracking to highlight differences in data collection practices related to category formation tasks in early development. Our second research objective was to investigate the role of words and/or arbitrary actions in the formation of visual categories for objects in infants at one year of age. We have recently submitted a manuscript on this topic to the Royal Society of Open Science journal, on May 12th, 2023.

Both projects are pre-registered on OSF and we aim to submit the second manuscript as soon as possible.

The authors confirm contribution to the paper as follows:

Study conception: Ricarda Bothe, Nivedita Mani

Study design: Ricarda Bothe, Léonie Trouillet, Birgit Elsner, Nivedita Mani

Data collection: Ricarda Bothe, Léonie Trouillet

Data analysis: Ricarda Bothe

Interpretation of results: Ricarda Bothe, Birgit Elsner, Nivedita Mani

Original draft manuscript preparation: Ricarda Bothe

Draft manuscript revision: Nivedita Mani

All authors reviewed the results and approved the latest version of the manuscript.

Abstract

Both words and gestures have been shown to influence object categorization, often even overriding perceptual similarities to cue category membership. However, gestures are often meaningful to infants while words are arbitrarily related to an object they refer to, more similar to arbitrary actions that can be performed on objects. In this study, we examine how words and arbitrary actions shape category formation. Across 3 conditions (word cue, action cue, word-action cue), we presented infants (N = 90) with 8 videos of single-category objects which vary in color and other perceptual features. The objects were either accompanied by a word and/or an action that is being performed on the object. Infants in the word and action condition showed a decrease in looking over the course of the familiarization phase indicating habituation to the category, but infants in the word-action-condition did not. At test, infants saw a novel object of the just-learned category and a novel object from another category side-by-side on the screen. There was some evidence for an advantage for words in shaping early object categorization, although we note that this was not robust across analyses.

1. Introduction

Words are accorded a special role in early development with the suggestion being that, unlike other auditory cues, such as tones, words lead infants to focus on the commonalities among these objects and, thereby, facilitate category formation (e.g., Althaus & Mareschal, 2014; Booth & Waxman, 2002; Fulkerson & Haaf, 2003; Fulkerson & Waxman, 2007; Plunkett et al., 2008). Alternatively, some studies suggest that infants' greater familiarity with words and the increased frequency of words in the input to the child, rather than specialized linguistic mechanisms, may underlie the role of words in shaping category formation. However, other cues in the input, such as non-verbal gestures and motions, are equally frequent and familiar to infants and have been shown to facilitate categorization in similar ways like words (e.g., Booth & Waxman, 2002; Sučević et al., 2021). Such studies fail to consider that novel words arbitrarily refer to objects, while gestures are linked in meaningful ways to the objects on which they are performed or to which they refer. What is missing is a comparison of how words and gestures modulate object categorization early in development when both cues are arbitrarily related to the objects. Against the background that arbitrary actions that can be performed on objects are more comparable to the arbitrariness of word-object associations than goal-directed gestures or motions, the current study will compare the role of words and arbitrary actions in shaping early object category formation. Our results suggest an advantage for words in shaping early object categorization, although we note that this was not robust across analyses. In what follows, we first outline the literature on how additional cues in the input, like words or actions, influence early object categorization and highlight our motivation for the current study.

Words – in contrast to other auditory cues like tones – have been shown to influence visual category formation from very early on in development. For example, Ferry and colleagues (2010) showed that when a label – but not a tone sequence – consistently accompanied objects from a particular category, even 3-month-olds show evidence of having formed a basic object category (e.g., all dinosaurs). Such findings are typically explained by suggesting that labels direct the infant's attention to perceptual commonalities among objects that share category membership (Althaus & Mareschal, 2014; Balaban & Waxman, 1997; Ferry et al., 2010; Fulkerson & Waxman, 2007; Sloutsky & Robinson, 2008; Waxman & Markow, 1995). Words may also go beyond merely triggering sensitivity to similarities in visual input during category learning, but may also introduce concepts that are not perceptually apparent onto a visual scene (e.g., Ferguson & Waxman, 2017; Waxman & Markow, 1995; Waxman & Braun, 2005, Waxman, 2003). For example, Fulkerson & Haaf (2003) reported that, by 9-months, infants form very basic object categories (e.g., horses, airplanes) when objects were accompanied by labels, sounds and in silence (e.g., no additional cue to visual object). However, when infants were presented with objects from more global, superordinate categories (e.g., animals, vehicles), only infants in the label condition showed category learning at test, suggesting that labels may be particularly powerful in shaping category formation, especially in the absence of category-related perceptual cues. Results from such studies have, therefore, been taken to suggest that language influences concept activation and category formation differently to non-linguistic information (Booth & Waxman, 2002; Fulkerson & Haaf, 2003; Fulkerson & Waxman, 2007; Lany et al., 2020; Lupyan & Thompson-Schill, 2012; Puccini & Liszkowski, 2012).

At the same time, some studies report no differences in the extent to which words and sounds impact early categorization success. For example, Robinson & Sloutsky (2007) familiarized 8- and 12-month-olds with a single-object category in silence or paired with either a word or sound and found increased attention to the objects during familiarization when they were accompanied by words and sounds, relative to when they were presented in silence. Differences in the extent to which infants attend to word-object mappings relative to sound-object mappings during familiarization may be driven by the familiarity of such cues. For example, words are more familiar to young infants relative to non-linguistic stimuli (Sloutsky & Robinson, 2008; Robinson & Sloutsky, 2007), because infants are exposed to language from early on. Thus, language may preferentially shape categorization due to infants' increased familiarity with linguistic cues. However, increased attention to the objects during familiarization in the word and sound condition was not mirrored in infants' category formation at test (see also Robinson & Sloutsky, 2004, 2007, 2008; Sloutsky & Napolitano, 2003; Sloutsky & Robinson, 2008). In this high-powered study (N = 162), infants showed improved category formation in the silent condition relative to when the objects were accompanied by words or sounds, suggesting that auditory cues impaired, rather than boosted category formation. Thus, while language influences infants' attention to objects, Robinson & Sloutsky (2007; and others cited above) find no evidence of an advantage for language in shaping early categorization. Nevertheless, if infants' increased attention to objects when accompanied by linguistic cues stems from their increased familiarity with language, then other familiar communicative cues, such as actions, should modulate infants' attention, and potentially categorization, to a similar extent.

Actions and gestures are frequent and extremely salient cues in the infant's immediate social environment. For instance, when caregivers interact with their infants, they often exaggerate and repeat movements with their hands (Koterba & Iverson, 2009) and use infant-directed actions such as pointing, giving, and showing (Capirci et al., 2005; Cheung et al., 2021) to increase infants' attention towards visual stimuli and the overall communicative input (Brand et al., 2002; Brand & Shallcross, 2008). Indeed, infants' ability to observe and understand actions as an important resource in early development was shown to operate independently from their experience with spoken language. Thus, both hearing and non-hearing infants aged 6-months looked longer towards infant-directed signing than adult-directed signing (e.g., Masataka, 1998) and 2- and 3-year-old hearing children use gestures that look similar to the signing of their same-aged deaf peers (Goldin-Meadow & Brentari, 2017).

By their first birthdays, infants actively encode functional cues – actions or movements that are performed on an object that results in an event or effect – to structure their visual input. For example, 11-month-olds showed evidence of categorizing objects with limited perceptual overlap (i.e., a single critical part was perceptually similar across objects), when objects were introduced with a functional cue during familiarization (i.e., placing the object on an apparatus allowed the experimenter to pull out a colored band), but not in absence of such functional cues (Träuble & Pauen, 2007). However, functional cues had no influence on infants' categorization of objects with increased perceptual overlap. Such findings suggest that form-function relations, just like object labels (c.f., Fulkerson & Haaf, 2003), can guide category formation when objects share limited perceptual overlap. Such functional knowledge continues to be an important asset throughout development and shows parallels with language by the second year of life, when children learn names for object functions and generalize theses names in keeping with the object's function (e.g., Kemler Nelson et al., 2000).

Studies directly comparing the influence of linguistic vs. functional cues in visual categorization suggest that functional cues may be more salient in shaping the formation of object categories. For instance, Booth & Waxman (2002) familiarized 14- and 18-month-olds with objects sharing perceptual overlap either accompanied by a word cue, a functional cue (e.g., jiggle or slide) or no cue. At test, infants saw a novel member of the familiarized category next to a novel object from an unseen category and were asked to touch/point to the object from the familiarized category, i.e., at test, infants were presented with the objects in the presence of the functional or word cue. Both age groups identified the target object at test in the functional cue condition. However, only 18-month-olds, but not 14-month-olds, identified the target object upon hearing its label, suggesting that functional cues – but not word cues – may trigger generalization of a novel instance of a familiar category early in development. In other words, infants successfully identify a novel member of a previously familiarized category when category members were presented alongside functional cues, but not word cues.

A more recent study by Sučević et al. (2021) investigated the effect of words and motions in visual category learning. Here, 10-month-olds were given the opportunity to explore objects from two novel categories that were either accompanied by a word, a motion (i.e., the objects moved up and down or side to side of their own volition) or no cue in a gaze-contingent design. Here too, infants showed improved category formation when the objects were accompanied by motions relative to words. Critically, category formation in this task was assessed by comparing infants' looking behavior to a novel object (here a composite of the two categories) relative to a prototype of the two trained categories, in the absence of additional word or motion cues. Additionally, gaze-contingent exploration showed that infants focused more on perceptually overlapping features across objects when they were presented with a cue (word or motion) relative to the no cue condition, especially when the objects were accompanied by motion cues. Thus, motion-, and potentially, word cues may drive infants' attention to the perceptual overlap between objects, thereby triggering generalization of objects with shared object features, leading infants to sort these objects into a common category (at least when accompanied by motion cues).

Results from such studies suggest that both word- and visual-dynamic cues support infants' encoding and recognition of category-specific properties. There is a crucial difference, however, in the association between the object and functional cues and object motions, on the one hand, and words, on the other hand. Functional cues, for instance, provide additional goal-directed information about causal form-function relations. Object motions, similarly, are initiated by the object and are related to the affordances of the object features. Words, on the other hand, are arbitrarily related to object features and may only cue visual categories once they have been successfully associated with object features in the first few trials. A more comparable examination of how infants use visual and linguistic information to cue category membership is afforded by examining the role of words and arbitrary actions in curing category membership.

Against this background, the current study examines how arbitrary word-object associations and arbitrary action-object associations cue object category membership. Across three conditions, we examined the extent to which infants categorize perceptually similar objects into a single category when objects were accompanied by either novel word cues, novel arbitrary actions being performed on the objects or both, word and arbitrary action cues. If category learning were equally influenced by accompanying words and actions, then infants should show similar evidence for category formation in both the word cue and the action cue condition. However, given that actions and gestures precede the emergence of words in early development (e.g., Bates et al., 1980; Capirci et al., 2005), we raise the possibility of an action-bias in category formation by 1 year of age. Moreover, of interest is the extent to which the presentation of simultaneous word and action cues shape category formation. On the one hand, words and actions often co-occur in the input with some accounts suggesting that the temporal alignment of words and actions in the input may boost learning of word-object associations (L. J. Gogate, 2010b; L. J. Gogate et al., 2006a). If multisensory redundancy were to similarly impact category formation, we would expect a boost in category formation in the word-action cue condition relative to the word cue and action cue condition. On the other hand, the simultaneous presentation of words and actions may detract attention from the perceptual similarities of the objects presented, potentially hindering category formation in the word-action cue condition.

Category formation is typically investigated by presenting infants with a series of perceptually overlapping objects from a single category during familiarization and then testing infants on their looking behavior to a novel member of the familiarized category and a member of a different as-yet novel category (Althaus & Mareschal, 2014; Plunkett et al., 2008; B. Younger, 1985). Longer looking times to the novel object from an unknown category relative to the novel object from the familiarized category at test is typically interpreted as generalization of the objects from the familiarized category and indexes object category formation (Bomba & Siqueland, 1983; see also Quinn, 1987). Some studies also examine infants' habituation to objects from the category during familiarization, with a decrease in looking time towards same-category objects over the course of the familiarization phase being interpreted as infants' habituation to the category (Althaus & Mareschal, 2014; Althaus & Westermann, 2016; Plunkett et al., 2008). We examined the proportion of looking time to the novel member of the familiar category relative to the novel member of the unknown category at test as one index of categorization, as well as the decrease in looking time to same-category objects during familiarization as an index of habituation to the category. In particular, we were interested in the extent to which arbitrary actions and words impact infants' object category formation, independently and in combination.

2. Methods

2.1. Participants

Infants were aged 1 year (N = 90, females = 47, Mage = 11.02 months, SDage = 1.07 months, age range = 10 – 14 months) and grew up in a German-monolingual environment (i.e., infants were exposed to German for > 80%; Byers-Heinlein et al., 2021). Infants were born at term or less than two weeks before term and parents reported no auditory or visual deficits with regard to their child's development and provided data for at least 50% of the training phase (i.e., look at the screen for at least half of a training trial, for four trials) as well as for the test phase (i.e., look at the screen for at least half of a test trial, for a single trial). Participation was rewarded with a book. The ethics committee of the University of Potsdam provided ethics approval for this project (No. 46/2018).

We pre-registered a sequential Bayesian analysis ("https://osf.io/a4rdt"), where we computed a Sequential Bayes Factor (SBF; Mani et al., 2021; Schönbrodt et al., 2017) following collection of data from individual children to examine the likelihood of the data under the alternative hypothesis, H_1

(infants look longer to the object from the familiar as opposed to the novel category), relative to the null hypothesis, H_0 (there is no difference in looking to the object from the familiar as opposed to the novel category). Here, if the BF exceeds a pre-specified minimum level of evidence (1/3 < BF < 3; i.e., moderate evidence for either hypothesis) for the alternative or null hypothesis, we would stop collecting data from additional children. We started computing the SBF following collection of data from a pre-specified initial sample of 20 children, in order to ensure that the initial sample size justified our belief in the effect (Schönbrodt et al., 2017; Mani et al., 2021). We planned, therefore, for our final sample size to be determined by the SBF values. However, we stopped data collection at 90 infants in order to optimize resource allocation.

2.2. Materials

We designed and produced two categories of soft toys to present infants with novel objects which were 20 cm tall and 9 cm wide, see Figure 1.

Figure 14

Objects used in the training and test phase

Note. Infants were familiarized with eight objects from a single category (i.e., either set A or set B) and then tested on their novelty preference as we presented them with one object from the just-familiarized category and one object from a never-seen-before category.

The final object sample (18 out of 24 soft toys) was chosen based on a visual similarity study with adults which took place online (N = 35, female = 20). Adults were presented with color photographs (12 soft toys of group A & 12 soft toys of group B) and asked to rate the similarity of object pairs within and across category items using a Likert-type scale from 1 (very dissimilar) to 6 (very similar).

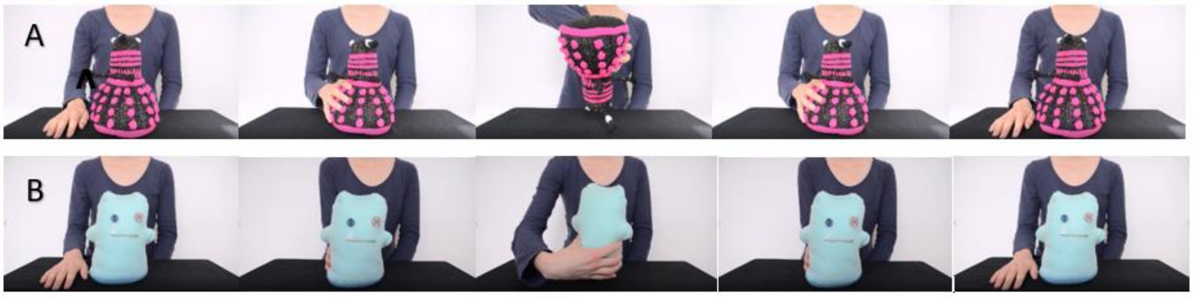
Similarity was rated via button press (1 – 6) on a computer keyboard. Pictures of object pairs and a similarity scale were shown on screen until participants pressed a button. Participants were provided with no auditory information during this study. Out of the 12 objects of category A and 12 objects of category B, we chose the nine objects in each category with the highest similarity scores relative to the objects in the other category. Objects included in the final sample for the infant study were all rated above 4 (out of 6), ranging from 4.17 to 5.03 in category A and from 4.14 to 5 in category B. We aggregated participant's ratings on object similarity within a category (i.e., each soft toy from category A with each other soft toy from category A) and between categories (i.e., each soft toy from category A with each soft toy from category B). Across participants, within-category ratings for soft toys from both categories were similar (category A: $M = 4.44$, $SD = 1.11$; category B: $M = 4.43$, $SD = 1.12$) and there was no difference in ratings between the two categories ($t = -0.06$, $df = 11.99$, $p = 0.95$). Similarity ratings were primarily driven by object shape (see Landau et al., 1992 for a shape bias in similarity detection), i.e., ratings of color-matched toys between categories was different from the within category ratings ($t = 2.41$, $SE = 1.14$, $p = 0.017$).

In total, 24 videos of the objects were created to form familiarization sets of eight stimuli in each category across each of the three conditions. In the *word cue condition*, infants saw the object standing still on a table centered to the screen, with the agent´s right hand on the table by the side of the object, palm facing down. The auditory stimuli for the training phase consisted of single presentations of the label i.e., "Oh, ein Tanu!" (Oh, a Tanu!) and were roughly two seconds long. Infants heard a label (i.e., "Tanu" or "Loeki", which are in keeping with German phonotactic constraints) produced by a female German-native speaker in infant-directed speech twice at 1000ms and 6000ms after video onset.

In the *action cue condition*, infants see the agent's hand performing an arbitrary action on each object and were provided with no auditory cue, see Figure 15. Within each video, action-object presentations were shown twice for each object. Videos began with the object standing in the middle of the screen and the agent's right hand to the right side of the object, palm placed down on the table. 1000ms after video onset the agent's hand starts to grasp the object and, across videos, either turns the object upside down and back up again or rotates the object along its axis, with the hand again by the side of the object facing down at the end of the video. The action is then repeated once more in the video, 2000ms after the completion of the first action (i.e., 6000ms after video onset). Each individual action was roughly 3 seconds long.

Figure 15

Visualizations of actions performed on the soft toys



Note. Infants saw eight objects from a single category either being turned upside down and back up again (A) or being rotated along its axis (B).

In the *word-action-object condition,* infants see the action being performed on each object while simultaneously hearing the label (i.e., “Tanu” or “Loeki”) in the presence of each object in a synchronized manner, i.e., onset of word-object and action-object presentations was temporally aligned (Gogate, 2010). Timing of presentation of stimuli was identical to the word cues and action cues described above.

At test, infants were presented with a photograph of an as yet unfamiliar object that overlapped perceptually with the objects presented during training alongside an image of an unfamiliar object that did not overlap perceptually with the objects presented during training, i.e., from the other category. Objects presented in the videos in the familiarization phase varied in color, while objects presented side-by-side at test matched in color.

2.3. Design

Infants participated in a novelty-preference task that consisted of a training phase and a subsequent test phase (i.e., individual training and test trials were 10s each). Infants were randomly assigned to one of the three conditions which differed with regard to the cue that was presented with the object during the familiarization phase only (word cue, action cue, word-action cue). Order of stimulus presentation during training and test was randomized for each participant. The labels and actions that co-occurred with specific categories were counterbalanced across infants. Thus, one group of infants heard the word “Tanu” or saw the upside-down action with objects from Category A, while another group of infants heard the word “Tanu” or saw the upside-down action with objects from Category B. The same was true for the other label and action.

Training phase.

The training phase consisted of 8 trials. In each trial, infants saw a video of a novel object from a single category. Object presentation was counterbalanced across participants such that half of the children saw objects from category set A during training and the other half saw objects from category set B.

Test phase.

In the test phase, infants saw photographs (854 x 480 pxl) of two objects side-by-side on the screen. Both trials were identical apart from left/right side presentation of test objects and lasted 10s each. Test objects were novel objects from category A and category B, such that the infant saw a photograph of an object sharing perceptual overlap with objects presented during training and one object from the other unfamiliar category. Infants were provided with no auditory stimuli during these windows. Longer looking times to the object from the unfamiliar category is typically interpreted as evidence for novelty-preference and taken to suggest that infants successfully categorized the objects presented during training.

2.4. Procedure

Infants sat either in a baby seat or on their caregiver’s lap at a distance of 60-55cm cm away from a TV screen (92 x 50cm) in a dimly lit experimental room. A remote eye tracker (Tobii X 120) placed underneath the TV screen recorded gaze data at 60 Hz and Tobii Pro Lab software will be used for stimulus presentation. Two loudspeakers placed above the TV screen presented auditory stimuli, while visual stimuli were presented on the TV screen. Two cameras located immediately above the TV screen provided online recordings of the infant and were used to keep track of the infant during the course of the experiment. We used a five-point grid calibration (a dot appearing in every corner and one in the center of the screen) in Tobii Pro Lab. The experiment was initiated following successful calibration and the first trial started when infants were fixating the screen.

3. Analyses and Results

We excluded data from additional infants due to eye-tracker calibration issues (*n* = 7), infant movement during the experimental procedure (*n* = 1), bilingualism (*n* = 2) and parental interference at test (*n* = 1).

3.1. Training data

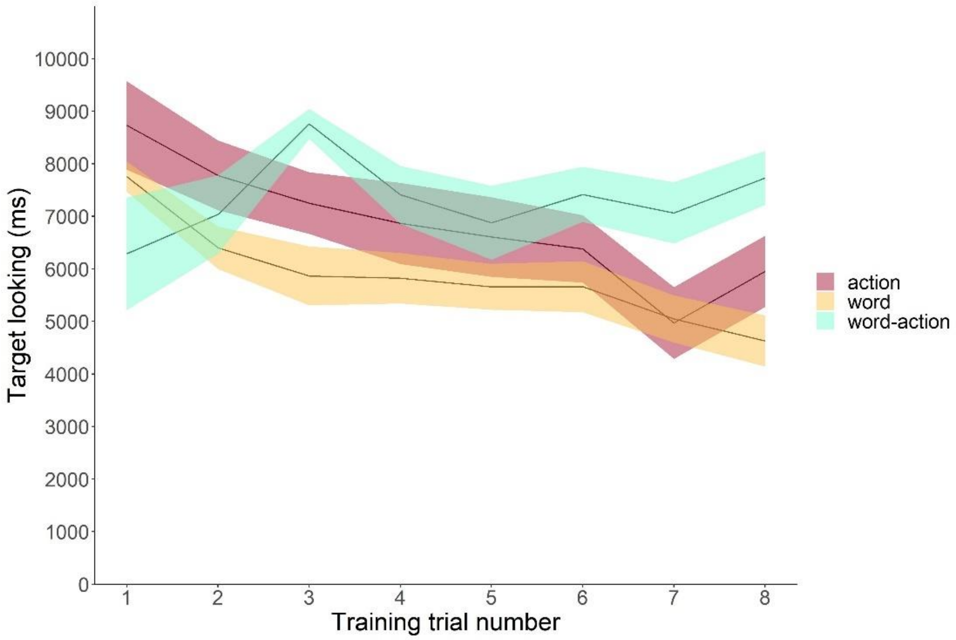
Altogether, the final sample provided us with 713 training trials (99.1% of all trials shown during the familiarization phase) while the average infant provided looking time data for 7.92 out of 8 possible training trials. During the familiarization phase, we examined whether there were differences in the amount of time infants spent looking at the screen when the objects were presented with accompanying words (word cue condition), actions (action cue condition) or words and actions (word-action cue condition). This allows us to identify potential differences in the extent to which infants habituated to the object category, i.e., showed a decrease in looking towards the objects with increasing number of trials, when they saw objects accompanied by a word-, action-, or word-action cue. The response was total looking time towards the object on screen in each trial.

To test the effect of *condition* on infant's looking during familiarization, we conducted a full-null model comparison. The full model was a generalized linear mixed effects model (GLMM; Baayen et al., 2008), using the function *lmer* of the package lme4 (version 1.1-23; Bates et al., 2015). The between-participant factor *condition* in interaction with the covariate *trial number* entered the fixed effect structure of the full model. In order to ease likelihood of model convergence and to enhance interpretability of estimate of coefficients, we z-transformed the covariate *trial number*. To allow for infant specific variation in looking behavior across the training trials for the individual infant, we add *infant ID* as a random intercepts effect and *trial number* on infant ID as a random slope (i.e., to minimize type I errors).

To avoid cryptic multiple testing (Forstmeier & Schielzeth, 2011), we compared this full model with the null model lacking *condition* and *trial number* as well as their interaction in the fixed effects parts but being identical to the full model in the random effects part. This comparison was significant, $\chi 2(4, N = 90) = 41.46$, $p < .001$, and the full model was of good stability, see Table 14. We then used drop1 analyses to identify which part of the fixed effect structure contributed to changes in the response, which revealed that infants' looking times were modulated by an interaction between *condition* and *trial number*, $p < .001$ (see Table 13). In other words, the extent to which infants decreased their looking towards object on screen over the course of familiarization phase, i.e., habituated to the presented category, differed across conditions, see Figure 16. As indicated by the significant interaction noted above, Figure 16 suggests that infants showed a decrease in their looking times towards the familiarization objects across trials in the word cue and action cue conditions, but not in the word-action cue condition.

Figure 16

Infant's looking trajectory towards objects during familiarization



As Table 13 suggests, while the decrease in looking time to the object on screen over trials did not differ between the action and the word condition, $p = .827$, as suggested by Figure 16, there was a significant difference in the extent to which infants habituated to the objects presented on screen across word and word-action trials, $p < .001$. However, despite the absence of a difference in the extent to which infants habituated in the word and action condition, infants spent longer overall looking at the object on screen in the action, $p < .001$, and the word action condition, $p < .001$, relative to the word condition.

Table 13

GLMM estimating the effect of condition, trial number and their interaction on infants' target looking

	Target looking during familiarization				
Predictors	*Estimates*	*std. Error*	*CI*	*Statistic*	*p*
(Intercept)	8.594	0.066	8.465 – 8.722	131.002	**<0.001**
condition [action]	0.447	0.093	0.265 – 0.628	4.818	**<0.001**
condition [word-action]	0.504	0.093	0.322 – 0.686	5.434	**<0.001**
z.trial.number	-0.168	0.039	-0.246 – -0.091	-4.272	**<0.001**
condition [action] * z.trial.number	-0.012	0.056	-0.121 – 0.097	-0.218	0.827
condition [word-action] * z.trial.number	0.177	0.056	0.068 – 0.287	3.181	**0.001**
Random Effects					
σ^2	0.16				
$\tau_{00\ id}$	0.11				
$\tau_{11\ id.z.trial.number}$	0.03				
$\rho_{01\ id}$	0.58				
ICC	0.46				
N_{id}	90				
Observations	711				
Marginal R^2 / Conditional R^2	0.194 / 0.561				

Table 14

Model stability full training GLMM

Predictor	*Estimate*	*min*	*max*
(Intercept)	-0.01	-0.10	0.04
conditionaction	1.78	1.72	1.88
conditionword-action	1.33	1.26	1.43
z.trial.number	-0.29	-0.31	-0.27
conditionaction:z.trial.number	-0.20	-0.23	-0.16
conditionword-action:z.trial.number	0.31	0.24	0.33
@(Intercept)	1.72	1.71	1.85
id@(Intercept)@NA	1.08	1.04	1.11
id@(Intercept)@z.trial.number	0.26	0.23	0.32
id@z.trial.number@NA	0.33	0.25	0.36

3.2. Test data

Infants provided us with a total of 175 test trials from 180 test trials shown (97.2%), with 85 infants delivering data for both test trials and 5 infants delivering data for a single test trial (word = 59 trials, action = 59 trials, word-action = 57 trials). In what follows, we first examine the extent to which infants' looking behavior, in particular, the proportion of time they spent looking at the unfamiliar object from the familiarized category relative to the object from the novel category, was modulated by the cues that accompanied presentation of these objects during familiarization, i.e., in the word cue, action cue and word-action cue condition.

We first report the results of our preregistered sequential Bayesian analyses comparing the proportion of target looking to chance in a Bayesian t-test separately for each condition. As reported above, we collected an initial sample of 20 children in each condition and then sequentially collected data from children in each condition and computed a sequential Bayes Factor upon adding individual children in each condition. The SBF in each condition is reported in Table 15. As Table 15 suggests, the SBF in the action and the word-action condition already crossed the threshold for the H_0 at 20 children (action cue condition, SBF=.306 and word-action cue condition, SBF=.276), while the SBF crossed the threshold for H_0 at 21 children in the word cue condition. We, however, continued testing children in all conditions to guard against a false negative (Mani et al., 2021). In the word-action condition, we stopped data collection at 30 children after we consistently found evidence for H_0 (SBF at 30 children = .198). SBFs for the action condition varied between .247 and .419. In other words, SBFs never neared the threshold for H_1 and consistently suggested anecdotal and, at 30 children, moderate evidence for H_0 in the action cue condition (SBF=.281). Thus, in both action and word-action conditions, we find consistent evidence for H_0 being true, given the data, i.e., that there was no significant difference in looking to the object from the familiar as opposed to the novel category.

In the word cue condition, while we found evidence for H_0 during testing, this pattern changed considerably with additional infants. With 30 children, we find anecdotal evidence for H_1 and stopped data collection. Thus, as far as the Bayesian analysis is concerned, our results are inconclusive with regards to whether we find a difference in looking to the object from the familiar as opposed to the novel category in the word condition.

Table 15

Development of the Bayes factor for each child per condition when n ≥ 20

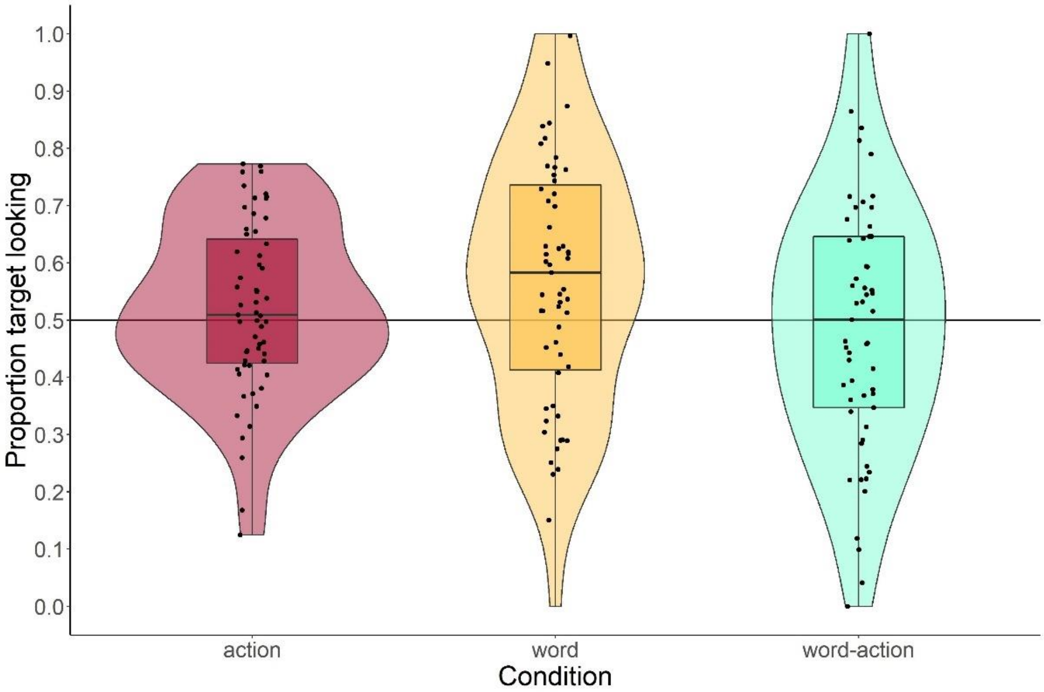
Condition	*SBF scale*	*BF development based on n*										
		20	*21*	*22*	*23*	*24*	*25*	*26*	*27*	*28*	*29*	*30*
action	medium	0.306	0.247	0.262	0.320	0.419	0.350	0.333	0.336	0.338	0.360	0.281
	wide	0.229	0.182	0.194	0.239	0.318	0.263	0.249	0.252	0.253	0.270	0.208
word	medium	0.336	0.304	0.407	0.331	0.409	0.533	0.460	0.620	0.868	0.931	1.091
	wide	0.253	0.227	0.309	0.248	0.310	0.409	0.350	0.479	0.680	0.731	0.860
word-action	medium	0.276	0.249	0.237	0.233	0.243	0.219	0.210	0.213	0.206	0.203	0.198
	wide	0.205	0.184	0.174	0.171	0.179	0.160	0.153	0.155	0.150	0.147	0.144

Next, we examine the effect of condition on infant's looking behavior at test using an GLMM analysis (Baayen et al., 2008). Here, the response was the proportion of target looking (PTL, i.e., looks to target divided by looks to target and distractor in a single test trial). The target here was defined as the object from the novel category, and the distractor was the other object from the familiarized category. The model was fitted assuming a beta error distribution (Bolker, 2008) and logit link function (McCullagh & Nelder, 1989), and analyses were conducted in R (version 4.1.2 or higher; R Core Team, 2022) using the functions 'glmmTMB' of the equally named package (version 1.1.1 or higher; Brooks et al., 2017).

Since the response included values being exactly 0 or exactly 1 (i.e., the infant spent all of the trial looking either at the target or all of the trial looking only at the distractor), we transformed the PTL response variable to allow for interval level assumptions of the beta distribution (Smithson & Verkuilen, 2006). With regard to the predictors, *condition*, the covariate test *trial number* (two levels: 1 & 2), as well as the difference between the looking time slope for each individual infant and the overall (fixed effect) slope during the familiarization phase entered the model as fixed effects. In order to increase likelihood of model convergence and to enhance interpretability of the estimate of coefficients, we z-transformed the covariates *trial number* at test and individual *habituation slope* during training. Since the dataset comprised only two observations per individual, the GLMM did not comprise any random slopes (Barr, 2013) but a random intercepts effect of *infant ID* and a random slope of *trial number* within infant ID.

Figure 17

Infants' proportion target looking at test per condition



To examine the contribution of *condition* on infant's looking preferences at test, we compared this full model to its respective null model (Forstmeier & Schielzeth, 2011), with the null model lacking *condition* in the fixed effect structure but being identical with regard to the random effect structure. This comparison test was based on a likelihood ratio test (Dobson, 2002) and was not statistically conclusive as to whether *condition* contributed to the variation in the response ($\chi 2(2, 90) = 5.393$, $p = .067$), see Figure 17. Drop1 analyses revealed that infants' *individual training slope* ($p = .853$) and test *trial number* ($p = .785$) did not account for variation within the response, nor did *condition* meet conventional threshold levels of statistical significance ($p = .067$). However, full model estimates revealed that proportion of target looking in the word condition (reference level) differed from chance, $p = .018$, suggesting that infants in the word condition looked significantly more at the object from the novel category relative to the object from the familiar category at test. Similar to the training model results, looking preferences in the word condition did not differ from those derived from the action condition, $p = .108$, but did, indeed, differ from the word-action condition, $p = .022$, see Table 16.

Table 16

Word reference: infants' PTL modelled based on condition, trial number and habituation slope

	Proportion target looking at test				
Predictors	*Estimates*	*std. Error*	*CI*	*Statistic*	*p*
(Intercept)	0.316	0.133	0.054 – 0.577	2.366	**0.018**
condition [action]	-0.260	0.162	-0.577 – 0.057	-1.609	0.108
condition [word-action]	-0.374	0.163	-0.694 – -0.053	-2.287	**0.022**
z.child.slope	-0.013	0.068	-0.145 – 0.120	-0.185	0.853
tr.trial.no	0.044	0.160	-0.270 – 0.357	0.273	0.785
Random Effects					
σ^2	-0.02				
$\tau_{00\ id}$	0.25				
$\tau_{11\ id.tr.trial.no}$	1.08				
$\rho_{01\ id}$	-0.80				
ICC	1.04				
N_{id}	89				
Observations	173				
Marginal R^2 / Conditional R^2	0.066 / 1.042				

Note. Word data was set to the reference level of condition.

We estimated model stability by omitting individuals individually, fitting the full model to each of the derived subsets, and then comparing the subset estimates to the estimates for the full data sets. This revealed the model being of good stability (see Table 17).

Table 17

Model stability full test GLMM

Predictor	*Estimate*	*min*	*max*
(Intercept)	0.32	0.27	0.36
conditionaction	-0.26	-0.30	-0.20
conditionword-action	-0.37	-0.48	-0.30
z.child.slope	-0.01	-0.08	0.02
tr.trial.no	0.04	-0.04	0.12
@(Intercept)	1.88	1.84	2.17
id@(Intercept)@NA	0.50	0.37	0.54
id@(Intercept)@tr.trial.no	-0.80	-0.87	-0.58
id@tr.trial.no@NA	1.04	0.92	1.06

While there was no statistical difference between the word and action estimates, we note that when the estimates were derived using the action condition as the reference level, the model indicated that the infants' looking behavior in the action group did not differ from chance, p = .681, see Table 18. Thus, there was no significant difference in infants' looking behavior to the object from the novel and familiar category. Neither was there a difference in looking behaviour in the action and the word-action condition, p = .490.

Table 18

Action reference: infants' PTL modelled based on condition, trial number and habituation slope

	Proportion target looking at test				
Predictors	*Estimates*	*std. Error*	*CI*	*Statistic*	*p*
(Intercept)	0.056	0.136	-0.210 – 0.322	0.411	0.681
condition [word]	0.260	0.162	-0.057 – 0.577	1.609	0.108
condition [word-action]	-0.114	0.165	-0.436 – 0.209	-0.690	0.490
z.child.slope	-0.013	0.068	-0.145 – 0.120	-0.185	0.853
tr.trial.no	0.044	0.160	-0.270 – 0.357	0.273	0.785
Random Effects					
σ^2	-0.02				
$\tau_{00\ id}$	0.25				
$\tau_{11\ id.tr.trial.no}$	1.08				
$\rho_{01\ id}$	-0.80				
ICC	1.04				
N_{id}	89				
Observations	173				
Marginal R^2 / Conditional R^2	0.066 / 1.042				

Note. Action data was set to the reference level of condition.

4. Discussion

In the above study, we examined how words and arbitrary actions individually and in combination impact infants' categorization of perceptually overlapping objects. To examine this, we familiarized infants with different objects belonging to a single perceptual category, accompanied by either word, action or word-action cues. Categorization in such tasks is typically indexed by infants habituating to the category during familiarization, i.e., showing reduced looking towards novel tokens of the category as familiarization continues, or looking more towards a novel object from a perceptually distinct category relative to a novel object from the same category at test. In the current study, infants in the word and action condition, but not in the word-action condition, habituated to the novel object category during familiarization, providing an index of category formation in the word and action, but not in the word-action condition. When tested on their looking behaviour to novel objects from the familiarized and a novel category at test, using a sequential Bayesian analysis, we found moderate evidence for the H_0 in the action and word-action condition, i.e., that infants did not look longer to novel objects from the novel category relative to the familiarized category when objects were accompanied by either actions or word-action cues during familiarization. We found anecdotal evidence for H_1 in the word condition, i.e., anecdotal evidence that infants looked longer to the novel object from the novel category relative to the familiarized category at test when objects were accompanied by a word-cue during familiarization. In contrast, a regression model suggested evidence for categorization in the word, but not in the action or word-action condition. Our results, therefore, hint towards an advantage for words in shaping early object categorization, although we underscore that this finding was not robust across analyses. Below, we discuss this potential word advantage in category formation over using arbitrary actions before evaluating the impact of accompanying actions on categorization.

Labels change the way infants see objects

The effects of accompanying word cues on object categorization are typically explained by suggesting that words increase sensitivity to the perceptual commonalities between objects and, therefore, promote the formation of a common category (e.g., Althaus & Mareschal, 2014; Booth & Waxman, 2002; Fulkerson & Haaf, 2003; Fulkerson & Waxman, 2007; S. Waxman & Markow, 1995; S. R. Waxman & Braun, 2005). This may be due, in part, to the fact that words are symbolic features and can be used to represent a wide range of objects and concepts (DeLoache, 2004).

A related account suggests that language cues (label-feedback hypothesis) may warp the perceptual space such that the accompanying labels may force objects closer in perceptual space, based on the statistical regularities of certain word-object pairings (e.g., Lupyan, 2012; Lupyan & Bergen, 2016; Lupyan & Thompson-Schill, 2012). After a couple of presentations of word-object pairings during familiarization in the current study, infants may find it easier to recognize other objects that share object features as belonging to the same category. Thus, words may play a particularly important role in allowing infants to generalize from specific instances of objects to broader categories. For example, Lupyan & Bergen (2016) illustrate with an example that it is harder to train a monkey than a human to climb a tree to collect coconuts, due to our ability to explain the task to humans using language. While a word can be used to refer to a single object, such as a coconut, this reference applies to all coconuts, although each coconut looks slightly different. So, the word *coconut* allows for irrelevant details of the individual coconut (like variation in size and color) to be abstracted over and serves as a stable reference to all coconuts. Other cues, such as actions, for instance, are specific to the individual demonstration of the action, such that reaching for a large coconut requires a hand gesture that is different from reaching for a small coconut. Actions may, therefore, not generalize over incidental features of objects, like the size of a coconut, and may not warp the perceptual space to aid detection of perceptual similarity in the same way that words do (Lupyan et al., 2007; Lupyan & Bergen, 2016).

Labels as features to categorization

Alternatively, the influence of words on early categorization has been explained by suggesting that infants may treat the word as an additional feature of the object, which in combination with other perceptual cues guides category formation (Althaus et al., 2020; Gliozzi et al., 2009; Plunkett et al., 2008). Thus, according to this feature-based account, it is not the case that words highlight the commonalities across the objects. Rather, the word is, here, merely a consistent common cue that accompanies members of the same category leading infants to form categories of objects that share similar perceptual features, and enlist these categories in learning the meanings of words.

Infants did not categorize objects when accompanied by arbitrary actions (and words)

While such an account would potentially predict a similar influence of other consistent accompanying cues on early categorization, here, we found limited evidence for infants' categorizing the objects when they were accompanied by actions or word-action cues. Indeed, the Bayesian analysis we report suggested moderate evidence that there was no difference in infants' looking behaviour to novel objects from the familiarized and a novel category, when these objects were accompanied by action or word-action cues during familiarization. Similarly, the regression model found no evidence to suggest infants discriminated between these objects in their looking behaviour at test. This finding stands in contrast to previous studies suggesting an influence of functional movements (Booth & Waxman, 2002) and object-initiated motions (Sucevic et al., 2021) on object categorization. In contrast to such cues that are more constrained by the affordances of the objects or are more intrinsic to the object (in the case of object-initiation motions), arbitrary actions are simply performed on the object and do not confer additional meaning to the action-object association. Thus, for example, infants could not relate an action goal to the action performance on objects within the category because arbitrary actions produced no visible or audible effects. Indeed, infants' expectation of an action effect or some kind of communicative intent, may have deterred attention to the perceptual features of the object and disrupted category formation. One explanation of these findings may, therefore, be that arbitrary actions do not have the same ability to reference meaning in a symbolic way (like words) and may be more limited in their ability to cue category membership.

Familiarization data: accompanying cues shape habituation differently

We did, however, find differences in the extent to which infants in the word, action and word-action conditions attended to objects during familiarization. In particular, during familiarization, infants in word and action groups showed a decrease in looking to the novel category members over the course of familiarization. Such a pattern of looking is typically interpreted as an index of habituation to the visual object category, and consequently, as evidence of category formation. Infants' looking behaviour during familiarization may, therefore, be taken to suggest that word and action cues similarly influence generalization of common object features, while the simultaneous presentation of word and action cues appears to disrupt such processes involved in category formation.

Importantly, we note that infants, in general, looked longer at objects longer when they were accompanied by an action cue compared to a word cue. Indeed, one possible explanation for the failure to find differences in looking behaviour to novel objects from the familiarized and the novel category at test in the action condition may lie in this increased attention to objects when they were accompanied by actions. In particular, the increased attention to the videos when actions were performed on the object relative to when the objects were merely accompanied by words may index the increased salience of actions on objects, detracting attention from the perceptual features of the objects presented. In what follows, we will discuss why infants may have attended longer to objects presented with an arbitrary action relative to when objects were presented with a word, and why such attentional differences to word and action stimuli may have affected category learning.

Attentional differences across modalities

In early development, dynamic hand movements such as grasping an object are extremely salient to infants. For instance, 5- and 9-month-olds looked longer at a novel object when the object was moved by an agent relative to when it was moved by an inanimate object (A. Woodward, 1998b). Thus, videos of objects being moved by an agent are likely to be more intriguing to infants than videos of stationary objects accompanied by words. Such differences in attention to objects during familiarization have also been shown to impact infants' performance at test in categorization tasks. For example, Robinson & Sloutsky (2007) showed that infants' attention to objects when they were presented silently versus when objects were accompanied by a tone or a word, affected whether infants showed evidence for category formation at test. In particular, word-object and sound-object presentations increased infants' visual attention to objects compared to silent presentations of objects. However, while infants in the silent condition showed evidence for categorization at test, this was not the case for infants in the word and tone conditions. Increased attention to objects accompanied by additional cues during object exploration may then affect infants' looking behavior at test and/or disrupt categorization during familiarization. There may, therefore, be a sweet spot for (examining) category learning that is mediated by attention to objects, such that either too little or too much attention to the objects within the category, detracts from categorization performance.

A further possibility is that infants found it more difficult to visually track an object when it was moving than when the object was stationary. Therefore, longer looking times for moving objects may reflect the time needed by infants to encode a moving object. Moving objects may disrupt infants' visual encoding to such an extent that it detracts from the typically presented novelty preference in such tasks at test. Additionally, infants who saw objects being moved may have expected the objects to be in motion at test as well. In this case, their expectations about the object from the familiar category would be violated because thus far, infants saw objects with perceptual overlap consistently presented with an action being performed on objects. They might even have been surprised to see both objects as stationary visuals on the screen, because up until the testing phase, objects were presented in motion, leading infants to expect objects from both the novel and the familiarized category to be in motion. Such violations of infants' expectations may have resulted in more balanced looking times to both objects, thereby, preventing them from showing novelty preference for the target object at test.

Alternatively, arbitrary actions performed on the object may have enabled infants to explore perceptual aspects of objects in depth, allowing them to better encode object detail compared to when the object was presented stationary. Indeed, during the action performance, objects were dynamically moved in space and presented from different angles, which revealed more perceptual details about the objects relative to when they were presented still. While the still presentation of objects in the word condition may have allowed for structured examination of the objects' perceptual commonalities, increased perceptual information about moving objects may have disrupted detection of perceptual overlap between objects. This, in turn, may lead to difficulties generalizing and conceptualizing these common object features. At the same time, we note that infants in the action condition habituated to the familiarized category, suggesting that they were sensitive to the fact that objects accompanied by a consistent action shared perceptual overlap. So why did infants in the action condition not show a novelty preference at test?

Modality preferences influence novelty preference at test

Aside from differences in attention to objects that were named, stationary or in motion, infants also display more general preferences for visual and auditory stimuli depending on the context in which these cues occur. For example, Emberson and colleagues (2019) showed that 8 to 10-month-olds differed in their preference for visual (e.g., smiling faces) and auditory (e.g., nonwords like *vot* or *meep*) stimuli, when the stimuli were novel or familiar. Specifically, infants in their study were first familiarized with six faces and six nonwords and then tested on their novelty preference for either another face, or another nonword, respectively. They found that infants displayed a novelty preference for auditory stimuli and familiarity preference for visual stimuli, i.e., they preferred familiar visual stimuli over a novel visual stimulus and a novel auditory stimulus over familiar auditory stimuli. This may have biased infants towards more of a novelty preference in the word condition, and a familiarity preference in the action condition, due to differences in infants' preference for visual and auditory stimuli. This would suggest that, while infants may have categorized the objects in both the word and the action condition, infants may be more likely to show category formation in the word relative to the action condition because category formation was indexed by novelty

preference. Furthermore, given that a novelty preference in infancy has been suggested to reflect more advanced stages of stimulus encoding (Hunter & Ames, 1988), our findings might be taken to suggest that infants show more advanced encoding of the objects in the word condition. The extent to which infants encoded objects during familiarization can thus depend on the modality of the stimulus infants become familiar with, on the one hand, and affect their novelty preference at test, on the other.

When word-object presentations were accompanied by actions, infants did not habituate to the category

Evidence for category formation at test was limited, when infants were familiarized with the objects and heard the label for this object and saw an action being performed on the object, similar to the action-only condition. However, unlike the action condition, infants in the word-action condition habituate to the object category during familiarization. The simultaneous presentation of words, actions and objects apparently led infants to pay close attention to the objects on screen with little to no drop in attention and, thus, no evidence of habituation to the familiar objects. Specifically, infants in the word-action condition looked at objects longer and more consistently than in the other two conditions, suggesting that encoding of the objects during familiarization may have been limited (see Emberson et al; 2019 and Hunter & Ames, 1988). In other words, simultaneous presentation of words and actions may have thrown off infants' attentional patterns to the shared object features as they explored visual objects. Similar to the action condition, infants' gaze behavior may be less structured when they observe objects being moved (and named) than when they see them stationary and hear their labels.

Even though actions and words can work together to enhance communication and learning (Gogate et al., 2006), multimodal presentations of temporally aligned word- action-object triads in the current study were detrimental to infants' category learning. Such results are in line with the literature showing that word-object learning is disrupted when words are accompanied by actions in early development (Bothe et al., under review; Eiteljoerge et al., 2019a; Puccini & Liszkowski, 2012). For instance, Bothe and colleagues (under review) familiarized 1- and 2-year-olds with novel, arbitrary word-action-object triads. Here, only 2-year-olds showed recognition of word-object and action-object associations, and that only when the word-object-action triads were presented synchronously. 1-year-olds, in contrast, showed no learning of either word-object or action-object associations when these stimuli were presented synchronously (or sequentially). This study extends findings from the associative learning literature by showing that combined word-action-object presentations similarly disrupt object categorization.

In conclusion, we found limited evidence for categorization when objects from a single perceptual category were accompanied by words, and potentially, actions, but not when they were accompanied by words and actions. Importantly, many of these findings were not consistent across analyses. Thus, the sequential Bayesian analysis found anecdotal evidence for categorization in the word condition and moderate evidence for absence of categorization (as indexed by looking behaviour at test) in the action and word-action condition. In contrast, the GLMM found evidence for categorization in the word condition but not in the action or word-action condition. The differences between the statistical approaches may result from the fact that categorization effects are relatively subtle in early development (Oakes & Ribar, 2005), which may not be accounted for in Bayesian analyses because there is no general way to represent uncertainty in the background knowledge of the data in this approach (Wang, 2004). We also found that infants habituated to the objects during familiarization in the word and the action condition, but not in the word-action condition. While this finding provides an index of categorization in the former two conditions, we note that, in general, infants looked longer at objects during familiarization in the action and the word-action condition relative to the word condition. The latter suggests that object encoding during familiarization may have been influenced by the objects being presented in motion. Taken together, our findings suggest that words, relative to actions and words and actions, may support categorization to an extent distinct from other cues that are equally familiar, albeit potentially more salient.

Chapter 4

Brief report on the relationship of word and category learning

Author statement

This brief report presents the initial written version of the rationale and design of an additional study being conducted as part of the PhD. Due to pandemic-related contact restrictions, the study was initially conducted online and then moved back to the laboratory when restrictions were allowed. Analysis of the category data in Chapter 3 revealed significant differences in looking time data between online data collection and eye-tracking methods. Therefore, we are currently conducting a replication of the entire study in the laboratory. Data collection is approximately 80% complete, with 64 of 80 infants already tested in the laboratory. The study design and rationale for conducting the study are presented separately in the thesis to provide a brief overview in the General Discussion.

The authors confirm contribution to the paper as follows:

Study conception: Ricarda Bothe, Nivedita Mani

Study design: Ricarda Bothe, Nivedita Mani

Data collection and analysis: Ricarda Bothe

Interpretation of results: Ricarda Bothe, Nivedita Mani

Original draft manuscript preparation: Ricarda Bothe

Draft manuscript revision: Nivedita Mani

Abstract

Early development involves both category learning and word learning, and the two seem to be related. Visual categorization of objects helps infants structure incoming information and guide processes related to object recognition and active vision. In word learning, a word is associated with a particular object or category. Previous studies suggest that words may support category formation more effectively than other cues, and a recent study has shown that category knowledge may also have an impact on infants' word learning success. It remains unclear whether these processes are independent or interdependent, and to what extent word learning influences category learning or vice versa. In this study, I examine whether infants learn words and object categories independently or whether they are interrelated. Two groups of 40 12-month-old infants participated in the study, completing a category learning task and a word learning task in different orders. In both groups, the first task was designed so that half of the infants had the opportunity to either learn word-object association or form object categories. Infants then completed a second task to test the influence of their learning from the first task. If the two processes are independent, then the one group of infants who saw the category task first should learn the object categories even if they did not learn a word for an object in the first task and vice versa. However, if the two processes are linked, then the infant's learning success in each task will depend on whether they learned in the first task. The results of this study will provide further insight into the relationship between word learning and category formation in early development.

1. Introduction

Category learning and word learning appear to occur early in development (e.g., Juszyk, 2003; Waxman 2002, Smith et al. 2001, Gopnik & Nazzi, 2003; Gelman & Koenig, 2001), suggesting that the processes underlying these learning tasks are interrelated. Visual categorization of objects, on the one hand, involves mentally grouping objects that are perceptually overlapping (e.g., Quinn & Eimas, 1996) or named the same (e.g., Plunkett et al., 2008). This reduces the processing cost of object encoding and helps infants structure incoming information from their environment (Althaus & Plunkett, 2016). Generalizing strategies underlying such category learning also allow infants to quickly structure incoming information from their immediate surroundings, which in turn helps guide processes involved in object recognition (Mash et al., 2007; Reynolds, 2015), active vision (Aslin et al., 2002; Spriet et al., 2020), and spatial orientation (McMurray & Aslin, 2004, Quinn, 2005). Word learning, on the other hand, involves associating a word with a particular object or category, which is done by quickly associating the word with the object representation (e.g., McMurray et al., 2012; Swingley, 2010). This process is initially fragile and can be reinforced by social cues such as joint attention with a caregiver or repetition in different contexts (e.g., Baldwin, 1995; Markman & Wachtel, 1988). However, it remains unclear to whether these processes simply co-occur or whether they are related, and if so, to what extent word learning influences category learning, or vice versa, in early development.

Numerous studies suggest that words can support category formation more effectively than other cues from the same auditory modality such as sounds (e.g., Balaban & Waxman, 1997; Ferry et al., 2010; Fulkerson & Haaf, 2003; Fulkerson & Waxman, 2007). In Chapter 3, I presented data showing that words also have an advantage in promoting category formation over other cues that were matched in familiarity and arbitrariness, such as arbitrary actions from the visual modality. These results suggest that words may have a significant impact on early cognitive development by facilitating category learning of objects. A recent study by Pomiechowska & Gliga (2019) showed that category knowledge may also have an impact on infants' word learning success. They conducted a study to investigate whether 12-month-old infants interpret object names as category names when only one object is named. During the study, infants were shown 18 pictures of objects from two different categories (staplers and coffee makers) in a blocked or randomized order, followed by naming one of the objects in a sentence. Infants were presented with objects from two different perceptual categories in either a blocked or a randomized order. In the blocked presentation, all objects from one category were shown sequentially before switching to the other category, whereas in the randomized presentation, objects from both categories were interspersed randomly. Infants who saw category objects in a blocked scenario during familiarization looked at the target object upon object naming, but infants who saw category objects randomly presented did not. In the test phase, infants were presented with a stapler and a coffee maker side by side while they heard the name of one of the two objects. Longer looking at the named object was taken as an index of successful object recognition upon naming and category knowledge. This was taken to suggest that preverbal knowledge about an object category may facilitate learning of the category name, as infants were able to extend the object name to an object with perceptual overlap when

Nonetheless, it is worth noting that the aforementioned study did not examine whether the infants actually learned the word, which is crucial since word extension can be challenging for infants at this developmental stage. Moreover, it remains unclear whether the presentation of objects within a category in a blocked manner is beneficial for learning or if randomizing the presentation would hinder learning. In order to gain further insight into the relationship between word learning and category formation, the author conducted a study involving 12-month-old infants. The main objective of the study was to investigate whether these cognitive processes are independent or interdependent, and to determine whether they have an influence on each other. The study therefore tested whether infants learn words and object categories independently from each other or if they are connected. Two groups of infants completed two tasks in different orders (i.e., either word-learning task first, henceforth Group A, or category learning task first, henceforth Group B). In both groups, the first task was designed so that half of the infants had the opportunity to learn (the word-object association or the object categories), while the other half were simply exposed to objects and words in an unsystematic way that made learning impossible. Then, the infants completed a second task to test their learning, i.e., the infants who participated in the word-learning task first participated in the category formation task second and vice versa. If the two processes underlying these types of learning are independent, then the infants in Group A (word-learning task prior to category task) should learn object categories even if they did not learn a word for an object in the first task. The infants in Group B (category task prior to word-learning task) should recognize objects when they hear a word even if they did not form object categories in the first task. However, if the two processes are connected, then the infants' learning success in each task will depend on whether they learned in the first task.

2. Methods

2.1. Participants

64 out of 80 1-year-old infants (range = 10-14 months) participated in the study so far. All participants grow up in a monolingual German-speaking environment and are exposed to their native language German at least 80% of the time (Byers-Heinlein et al., 2021). Families were recruited from the Babylab database and rewarded with a picture book for participation. Caregivers provided written informed consent and the ethics committee of the University of Goettingen granted ethical approval for this study (No. 46/2018).

2.2. Materials

2.2.1. Visual stimuli

Infants saw 22 static pictures of household objects (e.g., staplers, coffee makers, bottle and can openers) during the category and word learning task on a total of 20 training trials and 8 test trials. All images were presented against a white background. The images of the objects presented during the category training and word-object training were displayed sequentially and centered on the screen. The images of the objects at test (both the category test and the word-object recognition test) were presented side by side on the screen.

2.2.2. Speech stimuli

In the word learning task, infants were presented with a novel label in the presence of an object on the screen. Labels were in keeping with German phonotactic constraints (i.e., "Toma" and "Schufi") and were spoken by a female German native speaker. During word training, infants saw a single object on the screen and heard a single label five times in the presence of that object; first embedded in carrier phrases and then repeated again (e.g., "Das ist ein Toma! Siehst Du das Toma? Schau, ein Toma! Wo ist das Toma? Toma!"; this means "There is a Toma! Can you see the Toma? Look, a Toma! Where is the Toma? Toma!"). The training stimuli were 10 seconds long (the object was seen for 10 seconds) and the label was presented at 1.2 seconds, 3.3 seconds, 5.8 seconds, 7.9 seconds, and 9.2 seconds during the presentation of the object. At test, children saw the two objects side by side on the screen and heard one of the two labels embedded in a carrier sentence associated with one of the objects (e.g., "Schau, ein Toma!"; meaning "Look, a Toma!"), followed by the repetition of the same label twice ("Toma! Toma!"). The test stimuli were 7 seconds long (the object was seen for 7s) and the label was presented at 2.3s, 3.75s, and 5.16s. Word-object pairs were counterbalanced within the two groups. Specifically, all infants were presented with two novel objects, with half of the participants having one object labeled "Toma" and the other labeled "Schufi," while the other half were assigned the opposite labeling scheme.

2.3. Design

Participation lasted 3.65 minutes and included 28 trials (16 category familiarization trials, four category test trials, four word learning training trials, four word learning test trials). All trials began with a fixation cross centered on the screen for 1 second. Participants were randomly assigned to one of two conditions. Each condition (i.e., word learning task first or category learning task first) consisted of a word learning task and a category learning task. The two conditions were thus similar except for the order of task presentation, i.e., either the category task first and the word learning task second or vice versa (see Table 1). Infants participated in only one of the two conditions (i.e., the word learning task prior to the category task or vice versa), and each participation consisted of four phases: Category familiarization, category test (using the novelty preference paradigm), word-object training and word-object recognition test (using a preferential looking paradigm). In both conditions, word-object training was followed by a word-object recognition task, and category familiarization was followed by the novelty preference task.

In both conditions, the first task was systematically manipulated to elicit either learning or simple exposure to the word-object associations or category objects, so that it could be examined how learning success on one task affected performance on the subsequent task. That is, infants were either provided with the opportunity to learn or were simply exposed to the objects. See Table 19 for an overview. Thus, the study adopted a 2x2 design, with two variables considered as between-subjects factors: task order (word task first or category task first), learnability of the first task (see below).

Table 19

Study design to test the interplay between word and category learning

A: word learning task first	**Word-object task** →		**Category Task**	
Does word knowledge enhance category performance?	**Training**	**Test**	**Training**	**Test**
	Consistent	*Learnable*	*Blocked*	*Learnable*
	Inconsistent	*Not learnable (exposure)*	*Blocked*	*Learnable*

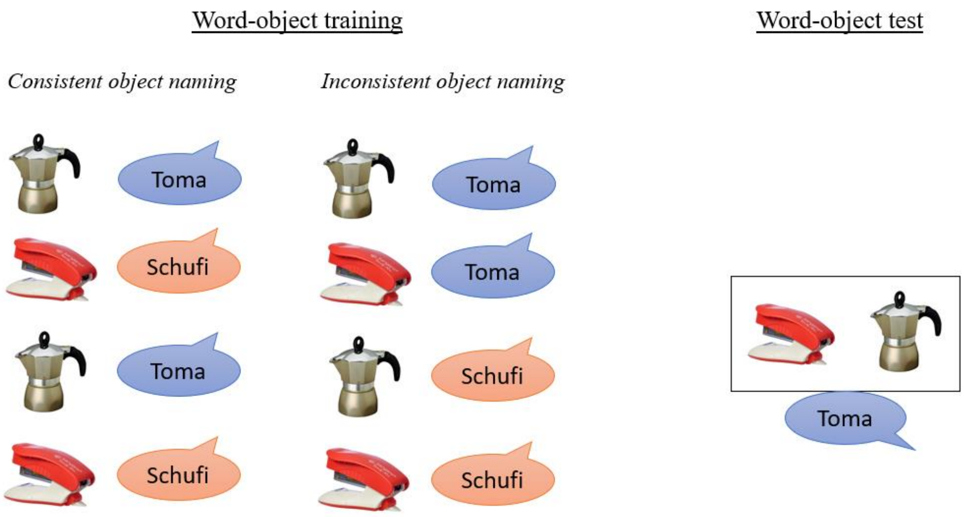
B: category task first	**Category Task** →		**Word-object task**	
Does category knowledge enhance word learning?	**Training**	**Test**	**Training**	**Test**
	Blocked	*Learnable*	*Consistent*	*Learnable*
	Random	*Presumably not learnable (exposure)*	*Consistent*	*Learnable*

Order of trials →

The term "learnability of the first task" refers to whether or not infants were able to learn from the input provided during the training phase. For the infants who saw the word task first, this means that the infants were able to learn word-object associations in this task when the objects were named consistently, or they were exposed only to words and objects without consistent object naming which made learning of the associations impossible. Specifically, infants who participated first in the word task (A) were introduced to two novel objects and two novel names, either consistently or inconsistently. That is, infants in the consistent group were presented with the same word-object pairs twice. Infants in the inconsistent group heard one name for a given object and then a different name for the same object on another trial. The latter made it impossible for infants to show learning for a word at test because each object was given the same two names, so infants in this group were exposed only to the objects and words, see Figure 21. In the subsequent category formation task, all infants in this condition (A) saw 16 objects from 2 perceptual categories (8 staplers and 8 coffee makers) presented in a blocked way, meaning that all staplers were presented sequentially before or after all coffee makers were presented sequentially.

Figure 18

Consistent and inconsistent naming groups in the word-first condition (A)

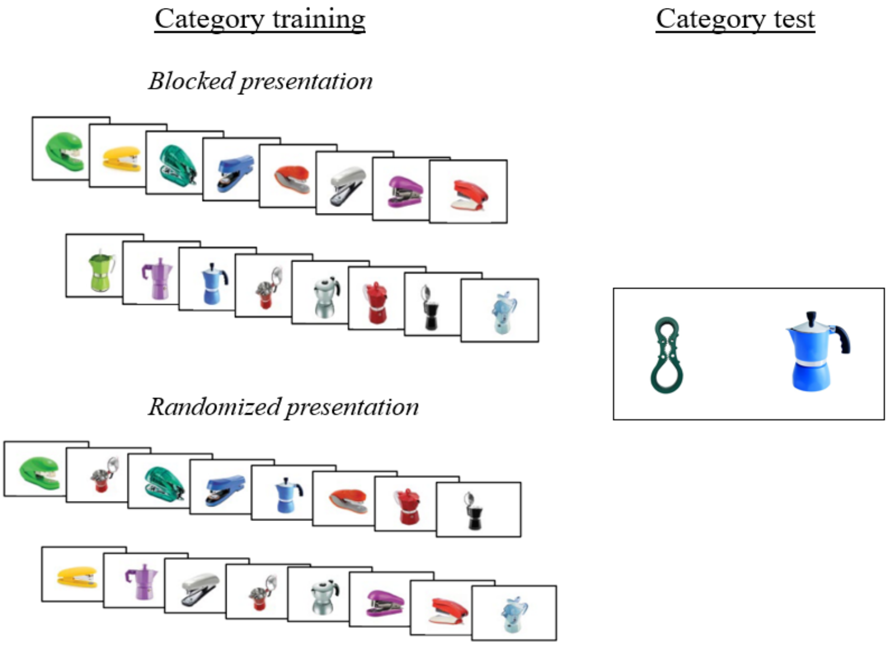


Note. All infants saw the same test images. For infants in the consistent group, the target object at test was the coffee maker as suggested by the word-object presentations during familiarization. There was no target object at test following inconsistent word-object presentations.

For the infants who saw the category task first, this means that the infants were given the opportunity to form object categories in this task when the objects within a category were presented systematically, or they were exposed to objects randomly which made learning of the associations impossible. Specifically, infants who participated first in the category task (B) saw 16 objects from 2 perceptual categories (8 staplers and 8 coffee makers), either blocked within the categories or randomly. That is, infants in the blocked group saw all staplers sequentially before or after all coffee makers were presented sequentially. The order in which the categories were presented in this group was counterbalanced. Infants in the random group saw all objects from both categories displayed randomly one after the other. In the latter condition, it was very difficult for infants to show learning in form of novelty preference for an out-of-category object given that objects with perceptual overlap were not introduced in a systematic way (Pomiechowska & Gliga, 2019), so infants in this group were mainly exposed to the objects, see Figure 22. In the subsequent word learning task, all infants in this condition (B) saw 2 objects, each named consistently the same way.

Figure 19

Blocked and random category groups in the category-first condition (B)



Note. The stimuli were identical to those presented in Pomiechowska & Gliga (2019) to allow for replication of the results.

The second task was always learnable. That is, all infants in Group A who participated first in the word task saw all objects in a blocked manner in the subsequent category formation task, that is, they saw all objects from one category (e.g., all staplers) followed by all objects from the other category (e.g., all coffee makers). The order in which the categories were presented was again counterbalanced, i.e., half of the infants saw the staplers first and then the coffee makers, and the other half saw the object categories in reverse order. Infants in Group B, who participated first in the category task, experienced the word-object presentations in the subsequent word-learning task in a consistent manner, i.e., they heard the same name for a single object and a different name for a different object on all trials. The naming of objects was counterbalanced again, i.e., half of the children heard the name "Schufi" in the presence of a stapler and "Toma" in the presence of a coffee maker, and the other half in reverse order.

2.4. Procedure

The infants in the study were seated either in a baby seat or on the lap of their caregiver, 60-55 cm from a TV screen measuring 92 x 50 cm in a dimly lit experimental room. Gaze data were recorded at a frequency of 60 Hz by a remote eye-tracker (Tobii X 120) located below the TV screen. Stimulus presentation was performed using Tobii Pro Lab software. Auditory stimuli were presented through two speakers located above the TV screen, while visual stimuli were presented on the TV screen. Two cameras positioned above the TV screen were used to capture online footage of the infant and track its movements throughout the experiment. Prior to the start of the experiment, a calibration was performed in the Tobii Pro Lab using a five-point grid with one point displayed in each corner and in the center of the screen. The experiment began as soon as the calibration was successfully completed. The first experiment began as soon as the infant had fixed the screen.

2.5. Analysis plan

The objective of our analysis is to assess the performance of infants on the second task, and examine whether their learning success was influenced by the condition of the first task (i.e., whether they had the opportunity to learn in the first task or not). We hence aim to explore the potential impact of the order of tasks on infants' ability to learn novel information in subsequent task.

2.5.1. Training trials

During the familiarization phase of the study, we conducted an analysis to determine whether the amount of time infants spent looking at objects on the screen differed as a function of their opportunity to learn on the first task (i.e., inconsistent vs. consistent word-object presentations in Group A and blocked vs. randomized object category presentation in Groups B). The purpose of this analysis was to determine possible differences in the infants' attention to the objects that might affect their learning in subsequent tasks. In Group A, where we investigated the impact of word-object knowledge on subsequent category formation, we aimed to identify any differences in the infants' habituation to the object category. This involved examining whether there were changes in the duration of time that the infants spent looking at the objects on screen over the course of the trials, as a measure of their level of habituation to the objects. Here, habituation refers to the fact that infants gradually decrease their looking at the in-category objects over the course of the familiarization phase. This process reflects the infants' increased familiarity with the objects, which leads to a decrease in their attention towards them (e.g., Althaus & Westermann, 2016). Similarly, in Group B, where we explore the effect of category knowledge on subsequent word-object learning, we aimed to identify differences in the infants' attention towards the word-object associations when presented consistently or inconsistently. This was again achieved by analyzing the total duration of time that the infants spent looking at the object on the screen during each trial.

To examine the effect of *learnability* in the first task on infants' looking behavior during familiarization in the second task, we conducted two model comparisons: one for category familiarization in Group A and one for word-object familiarization in Group B. We used generalized linear mixed effects models (henceforth GLMM; Baayen et al., 2008) for these comparisons, with the full models including the predictor *learnability* as a critical factor in interaction with *trial number*. The analyses will be conducted in R (version 4.3.0 or higher; R Core Team, 2022) using the function *lmer* of the package lme4 (version 1.1-32; Bates et al., 2015). The response will be target looking (TL) for each training trial per infant. In the word learning task, the familiarization phase consisted of four trials, whereas in the category formation task it consisted of 16 trials. To improve the interpretability of the coefficient estimates and to increase the likelihood of model convergence, we z-transformed the *trial number*. To account for individual differences in infants' looking behavior across training trials, we included infant ID as a random intercepts effect and trial number at ID as a random slope. To minimize the risk of type I errors and to avoid cryptic multiple testing, we compared the full model to a null model that lacked the predictors *learnability* and *trial number* and their interaction in the fixed effects parts but was identical to the full model in the random effects part. These comparisons allowed us to assess the influence of *learnability* in the first task on infants' looking behavior over the course of the familiarization phase in the second task in each *task order* group (i.e., A & B).

2.5.2. Test trials

We aim to investigate how the *task order* and *learnability* in the first task affect infants' looking behavior during a test in the second task. We will again use a GLMM (Baayen et al., 2008) to examine this relationship. Thus, *task order, learnability* as well as their interaction enter the model as critical predictors in the fixed effects structure. We will also include the covariate *trial number* (i.e., 4 trials in each group) as well as the difference between the *looking time slope* for each individual infant and the overall (fixed effect) slope during the familiarization phase. We will measure the proportion of target looking (PTL), which refers to infants' amount of looking at the target object compared to the amount of looking at both the target and distractor objects on screen. In Group A, we explore how word-object knowledge influences category formation. The target object in this group will be the object from the novel category, and the distractor object will be the object from the familiarized category. In Group B, we explore how category knowledge affects word-object learning. The target object in this group will be the object that infants previously associated with a specific word, and the distractor object will be the object that had a different name during the familiarization phase.

In order to increase likelihood of model convergence and to enhance interpretability of the estimate of coefficients, we will z-transformed the covariates *trial number* at test and individual *habituation slope* from the training phase. To account for individual differences in infants' looking behavior across training trials, we included infant ID as a random intercepts effect and trial number at ID as a random slope. We will compare this full model to a subsequent null model that lacked the predictors *task order, learnability* and their interaction in the fixed effects parts but was otherwise identical to the full model in the fixed and random effects part. This full-null model comparison will allow us to assess the influence of *task order* and *learnability* in the first task on infants' target looking at test in the second task.

3. Outlook

This study is motivated by previous literature suggesting that words can facilitate the formation of object categories and vice versa in early development. In Chapter 3, I examined the influence of words and arbitrary actions on category formation, both individually and in combination. Our findings from this study partially supported the hypothesis that words may have a greater influence on category formation than arbitrary actions. Building on these findings, we aim to investigate how infants' knowledge of words and objects affects their ability to learn about different object categories and vice versa. Specifically, we investigate how word-object knowledge and category knowledge are related and how they influence each other. To this end, we will use objects in a word learning task that belong to the same categories as the objects used in the category learning task. In this way, we can investigate the relationship between infants' ability to associate words with objects and their ability to categorize objects. Our goal is thus to better understand how words can help form categories and how the formation of categories can in turn influence word learning.

The data from this study will therefore allow to make assumptions on what role prior knowledge or expectations about object names or object categories may play in these mutual processes. If there is an interdependence between word learning and category formation, then infants who learn a word for an object should be more likely to form a subsequent object category relative to when they do not learn the word for that object. Similarly, infants who form a category of objects may be better able to learn words associated with another member of this object category. However, if these processes are independent of one another, infants should be form a subsequent object category whether they learned the word for that object or not. In this case, infants should associate words with an object regardless of whether they have formed a category. By examining these relationships, we hope to elucidate the mechanisms underlying these processes and gain a better understanding of how words and categories are linked in infants' early learning experiences.

Chapter 5

General Discussion

Words have been hypothesized to play a special role in early development because, unlike other auditory cues such as sounds, consistent labeling of objects pushes infants to recognize commonalities between them. However, words are more familiar to infants than sounds because words are part of the input that infants receive from the beginning. Actions are as common and familiar to infants as words because they frequently occur along with words in the infant's social learning environment. The present thesis investigates the influence of words and actions on early learning processes with the aim of comparing these cues in early cognition. Specifically, the goal was to determine whether words hold a special status in early development or whether other modal cues that share comparable levels of familiarity and input frequency can replicate the effect of words. The research approach adopted here thus provides an opportunity to understand the mechanisms of word and action processing in early learning and to advance our knowledge of how these processes shape the trajectory of cognitive development in the earliest stages of life. In Chapter 2 and 3, I presented data on the influence of word and action cues on infants' learning in two tasks, namely, a multisensory associative learning task in which both a word and an arbitrary action were presented with an object, and a visual category learning task in which either a word and/or an arbitrary action were presented with objects of a basic category. In both tasks, the actions were adapted to the arbitrariness of the word-object mappings and were therefore ideal candidates for comparison with words. The thesis hence investigates the role of words and actions in early learning processes, aiming to understand whether words hold a special status in early development or whether other cues that are equally familiar and arbitrary in their relationship to objects can have a similar effect. In what follows, I will briefly summarize the findings of each study and present them in the context of word and action effects.

Caregivers typically synchronize their gestures with object naming, resulting in a presentation of words, actions, and objects that provide redundant information for successful multisensory learning (e.g., Gogate, 2020; Maganti & Gogate, 2013). Previous studies with infants in laboratory settings have shown that the co-occurrence of actions can interfere with word learning when words and actions are not presented synchronously in the same visual scene with an object (e.g., Eiteljoerge et al., 2019; Puccini & Liszkowski, 2012). Thus, the lack of multisensory learning reported in the literature may be due to infants not benefiting from intersensory redundancy when words, actions, and objects are not presented synchronously. We therefore reexamined multisensory learning of words and actions for objects in the laboratory and manipulated the temporal dynamics underlying these presentations so that words and actions were shown either synchronously or sequentially (i.e., one after the other). As suggested by preferential looking towards the target object at test, we found that 2-year-olds benefited from the synchronous presentation of words and actions, such that they learned both words and actions for objects when these cues were first introduced synchronously but not when they were shown sequentially. In the associative learning study in Chapter 2, we report that the results of the preferential looking paradigm were not confirmed by the results of pupillometry in a subsequent expectation violation paradigm. In the latter paradigm, both match (initial word-action-object triads) and mismatch (modified word-action-object triads) trials were presented. One possible explanation for the lack of convergence between the two paradigms is that the violation of expectation task did not adequately capture the learning that occurred during the training phase prior to the preferential looking task due to a temporal delay between the learning event and the subsequent assessment of pupillary data. Alternatively, the mismatch trials in the expectation violation paradigm were more difficult to detect than the object recognition ability assessed in the preferential looking paradigm.

Infants at 1 year of age did not show learning of word-object or action-object associations in any of the time conditions or tasks. The fact that these associations were not learned in any of the presentation conditions (synchronous vs. sequential) may suggest that multisensory processing at 1 year comes at an additional cost or that the laboratory task did not provide ideal learning conditions for infants to learn at all. Across ages, then, infants either learned words and actions for objects or they learned neither word or action object associations. These results suggests that similar mechanisms underlie the processing of words and actions in response to the temporal dynamics of multisensory input, even though our data does not allow to make assumptions about underlying mechanisms. However, we find evidence for the fact that words and actions associative with objects in early development in a similar way.

The data from the category formation task in Chapter 3 suggest a different result. Specifically, the task demonstrated that infants did not generalize similar-looking objects based on arbitrary action cues or a combination of words and action cues. This stands in contrast to previous findings which suggested that functional cues and object movements were capable of enabling infants to categorize objects in a manner akin to words. For example, Sucevic et al. (2021) showed that 10-month-old infants can use motions to identify similarities between objects that move in the same way, similar to when they hear a consistent name for an object. However, the looking time data in the category formation task in Chapter 3 do not support these earlier conclusions when arbitrary actions were used instead of functional cues or movements. Regression models showed evidence of an advantage of words over arbitrary actions in basic category formation, but this was not consistent across analyses. Bayesian analyses were conducted to investigate the hypothesis that infants show signs of category formation, reflected in a novelty preference for the object outside the category at test. The results of these analyses revealed anecdotal support for the alternative hypothesis in the word condition, meaning that infants who were presented with word cues and objects showed some evidence for category formation at test. Conversely, the analyses in both the action and word-action conditions revealed moderate evidence for the null hypothesis, suggesting that infants in these conditions showed no evidence of category formation as they looked equally long at both objects. In other words, only infants in the word condition showed some evidence for category formation, while infants in the action and word-action did not. Thus, we presented some evidence that in category formation, arbitrary actions and words do not influence the underlying processes in the same way at age 1.

The origin of the differences between associative learning and category formation, and the mechanisms underlying the processing of word and/or action cues in the context of associative learning and category formation remain subject of investigation. These two processes enable infants to structure visual input in their immediate environment and rely on a variety of mechanisms that emerge during early development. While associative learning refers to the process by which infants learn to associate one stimulus with another, such as a word and/or action with an object, category learning refers to the process by which infants learn to group similar objects into categories based on these cues. The purpose of the data collected is to examine the nature of these processes in the context of temporal synchronicity in associative learning and arbitrariness of word and action cues in category learning to provide insight into the source of the potential advantage of words over actions in category learning.

In the following sections, I will discuss the main implications of the temporal dynamics as a crucial predictor of infants' learning success in the first study, and the arbitrariness of words and actions in relation to object categories in the second study. I will also address certain limitations of the results, such as the lack of learning shown by 1-year-olds in study I or the inability to replicate the learning effects in 2-year-olds using pupillometry. In addition, I will discuss the theoretical underpinnings of each study's findings to provide context that will help answer the overarching research question of this thesis: Are words special in early cognitive development?

Study I: Words and actions in object associative learning

The association of an object with an additional cue such as a word or an action depends on the infant's ability to process these additional cues and to form associations between them in order to recognize them when they are presented again. Thus, associative learning may begin in the moment infants experience two stimuli together, e.g., words and/or actions with objects, and association learning solidifies when they are repeatedly exposed to the associations over time (e.g., McMurray et al., 2012). In Chapter 2, we tested infants learning of novel word-object and action-object associations based on word-action-object exposure, while words and actions were either presented synchronously or sequentially. In a first test phase, infants saw two objects from familiarization next to each other on the screen, followed by a prime phase were infants either saw an action being performed on a see-through object or they heard a word in absence of visual cues on the screen before the two objects from familiarization were shown again on screen. Infants' preferential looking to the target object in this latter recognition phase (i.e., the object that the action or word prime referred to) indicates that they successfully identified the object with which the action prime or word prime was previously associated and that they excluded the other object as a referent for that prime. 2-year-olds who saw the word-action-object triads in synchronous timing provided evidence for word-object and action-object learning, as they were more likely to look at the object previously associated with the word prime or action prime compared with the other object. In contrast, 2-year-olds who saw word-action-object presentations sequentially did not identify the target according to the action prime or the word prime, suggesting that infants did not map the different action and word cues to the object during the word-action-object demonstration. Nevertheless, alternative methods such as pupillometry in a violation of expectation paradigm failed to reproduce these results. Moreover, 1-year-olds did not show learning of either action-object or word-object associations, which I will discuss in more detail in a later section.

Thus, the synchronous presentation of multisensory cues in the laboratory enabled 2-year-olds to associate words and actions with objects, whereas sequential presentations did not. It is possible that associative learning was facilitated in the synchronous group because this condition better corresponds to infants' natural learning environment, in which more than 59% of all word-action-object demonstrations are synchronized by the caregiver at age 2 years (Gogate et al., 2000). Infants are therefore used to such timed presentations of novel cues, and caregivers reinforce this temporal condition in infants' social environments by actively adapting to their infants' developmental needs, i.e., younger infants are more reliant on synchronous presentations than older infants (Gogate et al., 2014; Maganti & Gogate, 2013). Caregivers thus create ideal learning conditions for their infant because synchronous presentations of multisensory cues make some information in the input redundant when these cues overlap in time. This can help the infant identify non-redundant information in the input (Gogate & Bahrick, 1998, 2001), perhaps because non-redundant information is simply a portion of the total information provided in the input and may therefore reduce the load of processing the entire information in the input.

Implications

Temporal dynamics in early learning are multidimensional

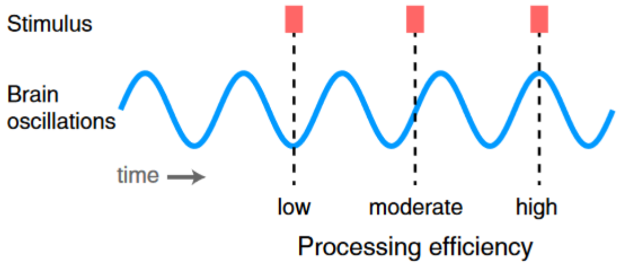
In the study reported in Chapter 2, I manipulated temporal synchronicity as a factor in word-action-object presentations, which has shown to improve learning of words and actions for objects by age two. Thus, 2-year-olds likely benefitted from intersensory redundancy when words, actions, and objects are not presented synchronously, which allowed them to recognize words and actions for objects in the subsequent test phase. In the following discussion, I will offer possible reasons as to why stimulus alignment could facilitate the detection of unique information in the input, leading to improved learning and recognition of associations between words and objects, as well as actions and objects. Based on findings indicating improved associative learning when words and actions were presented in temporal alignment with objects at 2 years of age, I propose that the mechanism underlying this intersensory redundancy is influenced by how stimuli are presented, since the timing of multisensory information presentation impacted the 2-year-olds' ability to detect and benefit from intersensory redundancy. Specifically, the impact of temporal factors on multisensory associative learning may extend to other temporal domains, including the oscillatory patterns of the infant's brain. When stimuli are presented in temporal alignment, it may create an ideal environment for learning and enhance the child's ability to recognize associations. Synchronization of oscillatory brain activity between the infant and caregiver may also occur in joint attention situations, potentially facilitating learning when temporally aligned stimuli are presented.

Neural coupling between timed input and brain activity

It could also be that the benefits of temporal synchronicity of multisensory cues in the input result from the neural patterns underlying associative learning. That is, the efficiency and robustness of stimulus encoding depends on the synchronization of the oscillatory activity of groups of neurons (Peelle & Davis, 2012). When neuronal activity in the sensory cortex is aligned with the temporal structure of the input, this creates an optimal learning state because ongoing neuronal oscillations match the temporal pattern of the stimuli (Lakatos et al., 2008; Schroeder & Lakatos, 2009), see Figure 18. Processing efficiency is increased when stimuli coincide with high-amplitude phases of the oscillatory signal, and efficiency is decreased when stimuli arrive at a time outside this window of high excitability, e.g., during low- or medium-amplitude phases (Peelle & Davis, 2012). Thus, infants' learning success is likely reflected in the temporal patterns in which the word, action, and object are presented to the infant. In this case, 2-year-olds who saw word-action-object presentations synchronously might therefore have benefited from timed cues because the stimulus presentation and brain activity were aligned, which facilitated processing of the input cues and thus enabled them to form robust associations. In contrast, the non-matched cues may have resulted in a temporal mismatch between neural activity and input that interrupted the processing of multisensory information, making association learning less efficient in the sequential presentation group.

Figure 20

Processing efficiency as a function of the temporal alignment between stimuli and oscillatory activity



Note. The illustration was taken from Peelle & Davis (2012).

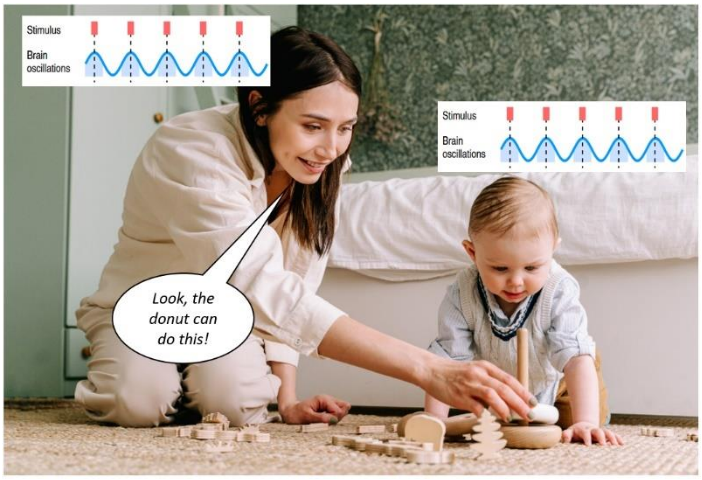
Interpersonal neural synchrony in early learning

By synchronizing their actions and word demonstrations for objects, caregivers can evoke this optimally tuned state between neural activity in the infant's brain and input cues during social interactions with their infant. These social interactions are complex processes that involve various cognitive factors such as attention to words, actions, and objects. When caregivers and infants engage in social interactions, their brains coordinate these processes to reduce uncertainty in the input and ensure successful interactions. This coordination of cognitive processes is also thought to be reflected in the synchronization of brain activity, which facilitates the learning process by allowing the infant to engage more effectively with input and process the social cues and information provided by others.

Thus, early associative learning involves not only the synchronous presentation of cues in input, but also interpersonal neural synchrony between the caregiver and the infant, i.e., the frequency of oscillatory activity between the two brains synchronizes and creates optimal learning opportunities within social interaction (e.g., Leong et al., 2017; Nguyen et al., 2022; Piazza et al., 2020). These synchronized patterns occur when caregiver and infant look into each other's eyes (dyadic synchrony) or focus their attention together on an object (triadic synchrony) during the interaction (e.g., Feldman, 2012). Figure 19 illustrates the temporal dynamics of such triadic interactions. Here, a mother and her infant naturally play with a toy to which they both give their attention.

Figure 21

Triadic synchrony in social interactions involving words, actions, and objects



Note. The illustration showing the interplay between stimulus timing and brain oscillations is from Peelle & Davis (2012).

The infant watches their mother stack a donut-shaped object on a wooden tower while the mother names the object in free speech. The moving of the object and the naming of the object are synchronized in time, allowing for temporal matching of the oscillatory signals during the word-action-object demonstrations. Joint attention to the multisensory event produced by the caregiver enables interpersonal neural synchrony between the oscillatory activity of the mother and the infant. Triadic synchrony involves coordination across multiple dimensions, including timing between caregiver and infant and between their brain activity and the input signals. Consequently, the success of early associative learning may depend not only on the temporal alignment of observed events such as words and actions with objects as tested in the study in Chapter 2, but also on the neural coupling of multimodal input and oscillations between the infant and the caregiver. This could explain why 1-year-olds did not learn word-object and action-object associations upon word-action-object, regardless of temporal alignment of the triads. This could be because additional levels of timing are required, such as triadic synchrony, which involves timing the oscillations of the caregiver and infant to support and recognize the learning of new associations.

Limitations

Exploring the absence of associative learning effects in 1-year-olds

When laboratory tasks with young infants lack social cues, learning may be compromised

The successful recognition of word-object and action-object associations in 2-year-olds may be due to temporal synchronicity effects that enable them to detect non-redundant information in the input. However, younger infants did not show learning under any of the temporal conditions, suggesting that 1-year-olds may rely more on social cues such as joint attention and joint oscillatory activity with their caregiver to associate cues in the laboratory task. Infants can benefit from social cues beyond simple word-action-object demonstrations, as these cues can more explicitly convey the intentions of interaction partners. Accompanying actions can be particularly helpful in enhancing the encoding of word-object associations by highlighting the object being referred to in the word. As infants begin to recognize the intentions behind others' actions, they also begin to understand the structure of intentional actions, which can help them to associate an action sequence with an object. For example, research by Baldwin and colleagues (2001) found that 10-month-old infants pay more attention to ongoing everyday actions when habituated to an action (e.g., a woman picking up a towel from the floor and placing it on a towel rack) if the action is interrupted in the middle, but not toward the end. This suggests that infants are able to extract intentional patterns based on the social context in which they occur and use this information to structure input and reduce ambiguity between actions and objects. So, infants can benefit from social cues that go beyond just word-action-object associations, as these cues can provide additional information about the intentions behind others' actions.

It is hence possible that the 1-year-olds in Chapter 2 did not learn the word-object and action-object associations from the multisensory word-action-object presentations due to the absence of additional factors that arise from social interactions with others. The laboratory tasks used in this thesis are typically conducted in rather sterile environments where infants look at a screen rather than learning in interaction with others. In their natural learning situations involving infant-caregiver interactions, infants are typically provided with multisensory information, including synchronized presentations of actions, words, and objects. However, these multisensory presentations may exceed infants' working memory capacities and attentional spans (e.g., Allen et al., 2006; Sweller & Sweller, 2006) in the laboratory, resulting in them being overwhelmed by the richness of the input. As a consequence, infants may not be able to identify non-redundant information in the multisensory input in laboratory settings even when word-action-object triads are presented in a timely aligned manner, and therefore not show learning of the word-object and action-object mappings.

Repeated exposure enables robust and slow associative learning

The fact that learning word-action-object associations in a laboratory setting can be challenging and beyond infants' attentional and memory capacities may have contributed to an insufficient amount of exposure during familiarization for infants. In the study reported in Chapter 2, infants saw both word-action-object combinations six times, and this number of exposures has been shown to be sufficient to promote learning of word-object associations (e.g., Ackermann et al. (2020) used five repetitions per word-object pair). Infants may need additional exposure to multisensory information because word-action-object triangles are more difficult for them to process and retain than simpler word-object mappings. However, increased exposure may improve infants' sensitivity to associations and allow for their learning over time (McMurray et al., 2012), based on the frequency of cues in their input (Saffran, 2003; Saffran et al., 1996). For example, the number of words infants are exposed to in the middle of the second year of life predicts how many words they will know 6 months later (Weisleder & Fernald, 2013) and infants' imitation of action sequences (e.g., putting a stick in a jar) in a picture book is enhanced when they are exposed to these sequences of actions (Simcock & DeLoache, 2008). Thus, given the fact that processing of multisensory input in the laboratory may be more demanding than in their natural environments where caregivers usually synchronize word-action-object demonstrations, infants may have benefited from additional presentations of the triads. Considering that processing multisensory input in a laboratory setting may be more challenging for infants compared to their natural environments where caregivers typically synchronize word-action-object demonstrations, it is likely that additional presentations of these triads would be beneficial because repetitions can increase the redundancy of information across sensory modalities, which may facilitate infants' attentional selection and perceptual learning (Bahrick & Lickliter, 2004b).

Uncovering the nuances of pupillometry analysis: Implications for study results

In this study, we observed that the analysis of the pupillometry data did not reveal any evidence for learning of the word and action object associations. The experiment involved presenting infants with word-object and action-object pairings, followed by testing their recognition of these pairings in a preferential looking task and a violation of expectation task. While 2-year-olds showed longer looking times at the target object following a word and action prime, there were no significant changes in pupil measures during the violation of expectation task. In the latter task, infants saw either word-object or action-object combinations, half of which were identical to the word-object and action-object associations from the familiarization trials (match trials) or were different combinations from those presented during familiarization (mismatch trials). If infants associated the word and action with the object during training, then the match trials at test should match their expectations of the associations, whereas the mismatch trials should violate their expectations of the word-object and action-object representations on the screen, which has been shown to be reflected in changes in pupil diameter (e.g., Jackson & Sirois, 2009; Pätzold & Liszkowski, 2019; Sirois & Jackson, 2012). Specifically, when expectations of the associations on the screen are not met, this has been shown to affect pupil size, such that it is increased for mismatch trials compared to match trials.

It is possible that I did not capture the effects of learning during this violation of expectation (VoE) task because the task was always presented last after the preferential looking time task. In fact, I could not balance the order of the tasks because the mismatch trials in the VoE paradigm may have interfered with learning from the familiarization, so the mismatch trials may have overridden the learning of the match trials, resulting in inaccurate recognition of the learned associations in the other task. It hence remains possible that the violation of expectation task would have indexed learning had it been presented immediately after the training task. Because of the fixed order, the learned associations may have been fleeting and difficult to recognize using the VoE task after a short delay of two to three minutes because the word-action-object training may have been short-lived, as infants' attention span and working memory capacity are limited (e.g., Allen et al., 2006).

Furthermore, infants may have difficulty perceiving changes in the word-object and action-object associations when confronted with mismatch trials, which may lead to different learning outcomes compared to matching trials. This suggests that matching and mismatching associations in early development may involve different learning processes or strategies, which could account for the absence of learning in this task. This idea is supported by the work of Ballem & Plunkett (2005), who argued that small changes in learned associations are difficult to detect in related paradigms such as the habituation and preferential looking paradigm (see also Swingley & Aslin, 2002). Specifically, when infants form associations between words, actions, and objects, small changes to these associations may not necessarily violate their expectations of the association. In such cases, infants would not show significant differences in their pupil dilations between match and mismatch trials, indicating that they do not perceive the changes as unexpected or novel. Thus, the ability to recognize matching and mismatching associations may depend on the complexity of multisensory mappings and infants' sensitivity to changes in these mappings.

Theoretical Framework

Analogous mechanisms in word and action learning for associative processes: Evidence from multiple domains

This suggests that words and actions are associated with objects in a comparable way and can be considered as features in the process of association. Thus, the data from Chapter 2 is likely to reflect a bottom-up approach to associative learning of words and objects. This assumption is supported by evidence from various domains, including animal behavior and artificial intelligence, which emphasize similarities among word-object and action-object associative learning.

In the animal kingdom, for example, dogs and monkeys show an impressive ability to learn associations between objects, actions, and words. Specifically, dogs can learn to associate hand gestures with objects or actions, such as fetching a ball or coming to their owner (Horowitz, 2009) and successfully retrieve an object in response to the spoken word previously associated with that object

(Kaminski et al., 2004) or retrieve the correct object when given a choice between two objects when naming an object (Pilley & Reid, 2011). Primates such as monkeys also show the ability to learn and use gestures for communication. For example, researchers have trained monkeys to use specific gestures to ask for food or indicate which object they want to interact with (Savage-Rumbaugh et al., 1993), and chimpanzees have been shown to retrieve objects based on spoken words (Terrace et al., 1979).

Recently, the use of artificial intelligence (AI) has highlighted the importance of associative learning. AI systems can use both associative learning and reinforcement learning to learn from data and make decisions based on acquired knowledge (e.g., Ghahramani, 2015). The core of learning here is to associate two entities based on their co-occurrence. For example, an AI model can predict language learning in humans through the use of computational models that simulate the process of language acquisition (e.g., Gliozzi et al., 2009). Infants' ability to learn associations between objects, actions, and words through associative learning could therefore be similar in animals and AI, underscoring the universality of associative learning as a fundamental learning process applicable to numerous research domains.

In short, associative learning of words, actions, and objects involves basic processes that infants typically master at an early age. Chapter 2 provides a detailed examination of the temporal dynamics underlying this learning and sheds light on why associative learning of words and objects may have been disrupted in previous studies involving actions. In this section, I also explored possible reasons why 1-year-olds showed no learning in the given task. For example, multisensory word-action-object learning can be particularly challenging for younger infants who may need additional social cues to facilitate their understanding of word-action-object associations. Additionally, the presentation of multiple cues may exceed their limited attentional and working memory capacities, which could be addressed by increasing the number of repetitions of the word-action-object associations during familiarization in future studies. Therefore, in addition to word and action cues and physical objects, other cognitive factors such as attention, memory, and social integration may enhance and reinforce learning. Furthermore, differences in methods and paradigms may affect the detection of subtle learning associations, as seen in the study in Chapter 2, where looking times, but not pupillometry, supported the results. It is possible that pupillometry did not capture the learning effects because this measure was only used in the second task which was further away from the initial learning phase and may have therefore failed to capture their transient learning effects. Alternatively, infants may have difficulty perceiving changes in word-object and action-object associations when confronted with mismatch trials, because small changes in these associations like a different action in the action-object presentations or a different word in the word-object presentations may not necessarily violate their expectations of these associations. Given the comparable effects of words and actions on associative processes throughout development and their parallel responses to similar temporal conditions, it is possible that these cues are regulated by analogous mechanisms in this type of learning. This claim is supported by evidence from diverse fields such as biology and technology, where word-object and action-object associative learning have been observed across species, including monkeys, dogs, and artificial intelligence.

Study II: Words and actions in object category formation

Such an analogous pattern of word-object and action-object associative learning in early development was not observed when words and actions were compared individually and in combination in a category formation task. In the category formation study in Chapter 3, I examined the influence of words and arbitrary actions on object categorization in 12-month-old infants. In the study, 90 infants were presented with eight videos of objects in a single category that were accompanied by either a word cue, an action cue, or both. Infants in the word and action conditions showed habituation to the category during the familiarization phase, but infants in the word-action condition did not. At test, there was some evidence that words had an advantage in early object categorization, although this was not consistent across analyzes. Next, I will explore implications of the study by first examining the role of arbitrariness in both the presentation of words and objects and the presentation of actions and objects in the context of categorization. I will offer explanations for why the presentation of arbitrary actions did not lead to similar effects in category formation as those of words. I will also highlight the importance of a control condition in which objects are presented to infants without additional cues, and outline how the location of data collection (online at home or in the lab using eye-tracking) affects the word effect observed in this study. I will then speculate on word effects in a broader discussion of the results, contrasting feature-based and inductive theories in the context in which words facilitate category formation.

Implications

The impact of arbitrariness: Word effects uncovered through comparison with arbitrary action cues

In category formation tasks, infants have the opportunity to generalize common object features and to group objects that share these features into a common category. This process does not necessarily require them to associate the name with the individual objects or the object category. For example, Hu, (2008) showed that 10-month-olds categorized objects based on a consistent word cue, but they did not show recognition for the category exemplars based on that word (for a summary of Hu’s 2008 study see section *Feature-based model* in the General Introduction). Thus, words can promote category formation of objects even when infants do not actually learn the word for the object category. Our results in Chapter 3 show some evidence that words as features are sufficient for the exclusivity of a category that infants form, and they are indicative of the relevant category distinction that the infant must make (Mareschal & Quinn, 2001). However, other cues that are consistent in terms of arbitrariness and familiarity, such as dynamic action sequences, do not lead infants to group objects with perceptual overlap and form a common representation of objects within the familiar category. Thus, the results of the category study suggest that there is something distinct about the role of words in early cognitive development. Specifically, when the arbitrariness of action cues was manipulated to make them more comparable to the arbitrariness of word cues, the uniqueness of words became more apparent. This uniqueness was not detected when words were compared to functional cues (Booth & Waxman, 2002) or self-initiated movements of the objects (Sučević et al., 2021) as described in the literature. Therefore, the category study in this thesis highlights the importance of considering the specific characteristics of cues when examining the role of words in early cognitive development. Infants, then, appear to rely less on actions than words for grouping objects together when the functional or self-initiated component of an action is removed. In what follows, I will provide a few potential reasons for why actions didn't have the same impact as words in basic category formation.

Arbitrary actions in category formation

Unlike words, actions tuned to the arbitrariness of word-object associations did not lead infants to form an object category. Seeing an object in motion might help infants encode the objects more fully as they are presented from different angles, but it could also make it harder for them to visually remember the object. For example, it is possible that infants had a harder time recognizing similarities between objects when they were observed in motion, because unlike in the word condition, in which objects were still on the table, infants in the action condition were unable to grasp object features in detail over an extended period of time. If infants in the action condition could not fully encode the individual objects because they had to track them in motion and from different angles, the similarities between the objects may have been less obvious than when objects were still.

At the same time, the moving objects allowed for a richer visual scene that was complex in terms of the details of the objects from different angles, e.g., they saw the back of the object, or the body turned upside down. Thus, the actions entailed high perceptual variability in the visual scene, which may have placed additional strain on the processing of similarity recognition of object features. It may have been more difficult for infants in the action condition to process and encode the object in motion than for those who saw the object motionless and named. Increased visual perceptual variability in the action and word-action conditions might have placed greater demands on processing than when visual perceptual variability was low in the presence of linguistic cues. Actions, then, may be more variable in a way that they represent commonalities of objects and therefore more difficult for the infant to recognize visually. Ultimately, the optimal level of variability and consistency is an important consideration for effective learning (Raviv et al., 2022), meaning that on the one hand, a lot of variability in what we see can make learning difficult, but on the other hand, some types of variability can actually be helpful in identifying similarities between objects.

Words, however, are less variable in the way they accompany objects because they introduce no additional visual information to the scene. This pattern was also reflected in the study in Chapter 3 in looking times during familiarization, when infants looked significantly longer at objects that were moved than when they were motionless. The perceptual richness of the visual scene thus seemed to make it very interesting for the infants to look at the objects on the screen, perhaps indicating novelty and thus less robust habituation in the presence of the action. In both the word and action conditions were habituated to the category by the end of the familiarization phase, but the total time spent looking was longer when the action was present. The increased visual interest in objects in the action condition compared to the word condition might therefore suggest that it was easier for infants to encode objects when they were named than when they were moved, so that infants looked longer at the action event to figure out what was going on or to make sense of the situation. Nevertheless, the infants showed successful habituation in both the word and action conditions, so it would have been expected that they would show a novelty preference for the object of the unfamiliar category at test as an index of successful generalization of the objects from the familiar category.

Infants prefer familiar visual cues but novel auditory cues

Although infants in both the word and action conditions habituated to the within-category objects, only infants in the word condition provided evidence of category formation as indicated by novelty preference for the unfamiliar category object. These differences may be explained by infants' preference for familiar visual cues over novel ones, but for novel over familiar auditory cues (Emberson et al., 2019, see also Cohen, 2004, for a discussion on familiarity preference in novelty preference paradigms). While familiarity with the word object may have caused infants to visually turn toward the unfamiliar category object, presentations of action objects may have caused infants to turn toward the familiar category object at test. Because novelty preference in infancy reflects an advanced stage of stimulus encoding (Hunter & Ames, 1988), it could be that infants in the word condition encoded named and motionless objects more thoroughly than infants in the conditions that contained actions.

Word-action-object presentations disrupt category formation

When words were added to the action-object presentations in the same category study, infants in the word-action condition showed high visual interest in the objects throughout familiarization and thus showed no signs of habituation at the end of the familiarization phase. It could be that these infants were constantly engaged with the visual stimuli on the screen because the combination of action and words made the experience particularly interesting or overwhelming to them, i.e., to an extent that they showed no signs of familiarity with the category. This could account for the processing difficulties in encoding the object, since the action and word may impose additional processing demands on object encoding compared to presenting an action or word with an object alone.

Similar to the 1-year-olds in the study in Chapter 2 who were presented with associations of words, actions, and objects either synchronously or sequentially but showed no recognition for any of the action or word cues at test, the 1-year-olds in the combined word-action condition in the category study in Chapter 3 showed no evidence of category learning at test. These results are consistent with research showing that when infants have the opportunity to perceive stimuli from different modalities simultaneously, they focus on a single modality (Lewkowicz, 1994; Lewkowicz, 1988; Sloutsky & Napolitano, 2003). Such studies report an advantage of auditory processing over the display of visual stimuli to the extent that auditory information overshadows visual stimuli. If the words in the word-action condition to some extent overshadowed the already complex presentations of objects in motion, it would be very difficult for infants to encode the multisensory information and to detect similarities between objects. Thus, the combination of word and action presentations may comprise infants' information processing systems to an extent that the infants do not show evidence for category formation at test.

It could be that 1-year-olds rely on the social context in which objects are presented for category learning, using word and action cues simultaneously. That is, isolating perceptually overlapping objects with words and actions from the social interaction in which these cues typically occur might reduce their ability to recognize object similarities and generalize them to a common representation. 1-year-olds, therefore, may not have been able to use strategies to process stimuli based on intersensory redundancy because the temporal synchrony of the word-action-object triads in the category task was insufficient for them to recognize these cues as markers of object similarities. Instead, they may have required that the word-action-object presentations be embedded in a social interaction at the moment of processing, so that the triadic synchrony between caregiver, infant, and the moving and named objects could assist them in focusing their attention on the objects that shared perceptual overlap to enable similarity detection. Alternatively, repetition of the word-action-object triad in different contexts may have further supported category learning, as this would have allowed infants to examine moving and named objects over time and in different situations, allowing them to more thoroughly evaluate similarities between physical objects.

Thus, when arbitrary actions accompany word-object associations in a category learning task, infants show no signs of habituation or category formation. If actions do indeed affect object encoding and visual novelty preference at test when performed either alone or in combination with words, then actions are likely to interfere with category formation. These results raise the question of whether words actually play a role in category learning, because it is possible that infants showed signs of categorization even when only the object was shown without the word, so words do not actually enhance categorization. In the next section, I will discuss the challenges of interpreting the effects of words on category formation and the importance of a control condition in this regard. In addition, I will explore the importance of the location of data collection in evaluating evidence for category formation based on words.

Limitations

Data in context: The importance of a silent control in object categorization studies

To fully understand the role of words in facilitating category formation, it is important to consider the extent to which infants can categorize objects without additional cues. To this end, an additional group of infants could be included in studies to evaluate the effect of simple object presentation without naming on infants' ability to categorize objects. Previous research has shown that even infants as young as 3 months old can categorize objects based on surface features, such as shape (e.g., Landau et al., 1992), and that 10-month-old infants can form perceptual categories of objects based solely on the shape of specific object features like the length of the neck (Plunkett et al., 2008). These findings suggest that infants have the ability to categorize objects based on perceptual features even without additional cues, but further research is needed to determine the specific role of words in facilitating category formation.

However, others have reported that infants have difficulty categorizing objects with perceptual overlap when there is no consistent word accompanying the objects. For example, Waxman & Braun

(2005) found that 12-month-old infants categorized animals with similar shapes only when these objects were presented with a consistent name ("Look, it's a keeto!"), but not when the objects were named differently or accompanied by a sentence without a novel word ("Look here!"). In these studies, the latter condition is often referred to as the "unsupervised" condition because it provides linguistic support but no novel word for the object. It is possible that this unsupervised condition confused infants because they may have expected a novel word in the presence of a novel object, so the unsupervised cues may have increased children's attention to object naming in a manner similar to actions, possibly leading to less robust habituation and impaired category formation (see also Plunkett et al., 2008 and Sloutsky & Napolitano, 2003, for a similar argument). Consequently, it is impossible to determine whether the observed category effect is solely attentional or driven by an association between object and label. Thus, without a control condition in which children see the objects in silence and without linguistic cues, there is little to no support for a word effect in category formation, because infants could categorize the objects based on their shape even without words. It is possible, therefore, that the evidence for category formation in the word condition in Chapter 3 reflects not the facilitation of object generalization, but the ability of infants to categorize objects based on similarities such as object shape, regardless of labeling.

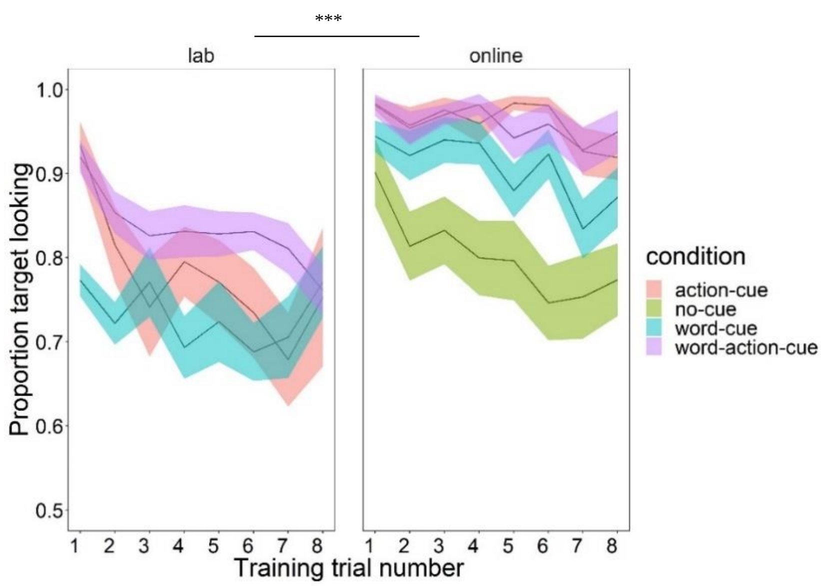
Exploring the impact of data collection location on infant perception studies

To ensure that the effects of words in category learning observed in Chapter 3 were not due to phenomena like the shape bias, a fourth condition is being tested where 1-year-olds view objects without any naming or movement during familiarization. The study was originally conducted online in 4 conditions (silent, word, action, word-action), but no evidence of category formation was found in any of the groups. The study was then moved to the laboratory when pandemic restrictions allowed, and eye-tracking data was collected instead. Unfortunately, these external factors prevented the complete replication of the study in a laboratory setting before the submission of this thesis. Therefore, I can only speculate about the effects of words on category learning in the following discussion. Given the differences in visual perception and stimulus attention that depend on the environment in which the data are collected, we decided to publish the results related to the word condition in a separate article. This decision is based on the fact that we only found an effect in the word condition in the laboratory setting, but not when participants were tested online using a laptop screen at home. Nevertheless, the comparison of category learning results between the online and laboratory settings highlights the subtle nuances in categorization effects that may arise when testing infants in different environments, such as at home on a laptop screen versus in a controlled laboratory setting.

What might be the reason for such location differences? We found differences in the way infants attend to objects when they were shown on a smaller laptop screen compared to the larger screen in the laboratory (see Figure 20). More specifically, infants who saw the objects along with a word and/or action cue on a laptop screen online looked at the objects longer on average than infants who saw the objects in the laboratory (see also Oakes et al., 1996, for task-dependent category formation). Thus, *where* infants saw the objects moved and/or named influenced the amount of time they spent looking at the objects during word-object and/or action-object familiarization, which in turn affected whether they formed an object category at test.

Figure 22

Looking behavior during object category familiarization when infants were tested online (Chapter 3)



Note. No data for the no-cue (control) condition in the laboratory because data collection is not yet complete.

The difference in how long infants look at things on laptop screens versus in the laboratory may show that they like looking at things on laptops more. This could be because it's harder to focus on small screens, and infants expect more interesting things on laptops. This means that if infants are used to watching exciting videos on laptops, they might expect the same level of excitement during experiments. Such factors can potentially impact an infant's level of engagement with an object, their level of interest, and the degree to which they encode the objects. Additionally, the absence of category formation during testing among all groups tested online may suggest that the encoding of familiar category objects was not very strong (see Hunter & Ames, 1988, for the source of this argument) when objects were displayed on a small screen relative to when objects were displayed in the laboratory. It could therefore be that the infants who participated online did not adequately encode the objects during familiarization in any of the conditions. Thus, inadequate encoding could be one of the reasons that even word presentation did not lead infants to categorize objects even though they showed evidence for categorization based on words in the laboratory.

To provide more conclusive evidence, it is important to include a silent and motionless control condition in the laboratory setting that does not contain additional cues, as a means of comparison. This control condition would allow a comparison between the effect of words on category formation and the formation of object categories without word cues. This comparison would be crucial because the only group of infants who showed category formation at test were those who saw objects in the laboratory while being exposed to a consistent word for each object. Interestingly, these infants also spent the least amount of time fixating on the objects during the habituation phase compared to all other infants tested both online and in the laboratory. It is likely, therefore, that the amount of time infants spent visually exploring the objects was directly reflected in their novelty preference for the out-of-category object at test shortly after the habituation. It is plausible that a comparable trend could occur in the control condition, in which infants are presented with objects in the laboratory without additional cues. This speculation is based on the fact that infants in the control group tested online paid significantly less attention to the objects than those exposed to additional word and/or action cues in the online condition.

Although the lack of a control group in the laboratory condition and the observed variations in looking times and category formation at different data collection sites impose some limitations, I will continue to argue for the word effect in category formation at age 1. To support this position, next I will examine the strategies infants may use to identify commonalities between objects based on a word in the context of bottom-up and top-down approaches.

Theoretical Framework

Category formation through labeling: features or induction?

The data presented in Chapter 3 suggests that words have a unique role in early cognitive development, as they appear to promote the formation of object categories more effectively than arbitrary actions. It remains unclear whether this advantage stems from feature-based processes in a bottom-up processing context or if words induce category understanding onto the visual scene in a top-down fashion. In what follows, I will examine feature-based strategies, which assume that words are features of objects similar to visual characteristics of the objects, and contrast these with inductive theories that assume that words change the way infants see category members in the visual scene. While the feature-based approach can be seen as a bottom-up strategy, as it begins with the exploration of individual features and builds up to the categorization of objects based on those features, inductive strategies involve forming categories based on higher-level concepts. This approach can be seen as a top-down strategy, as it starts with an overarching concept and applies it to the categorization of individual objects. Note that the data available from Chapter 3 do not allow us to make assumptions about the underlying mechanisms and theoretical framework that might play a role in producing word effects in category formation. Therefore, I will limit the following discussion to speculations about possible effects rather than drawing conclusions about potential mechanisms.

Feature-based processing as a bottom-up approach to category learning

From a feature-based perspective of information processing, infants might have used prototype-based (e.g., Quinn & Eimas, 1996) and exemplar-based (Oakes & Spalding, 1997) strategies to form categories of objects based on a word, but not based on an arbitrary action performance. However, since I did not test this directly, it is unclear what strategy the infants actually used, so I can only present assumptions and possible explanations for the results obtained.

One possibility is that infants have formed a category based on the word allowing them to represent a typical or representative exemplar of the members of that category. In Chapter 3, infants may have built a mental representation of the observed object and then updated that representation when they saw another object of the same category that was named the same. In this way, by the end of the habituation phase, infants may have built a representation of all objects in the category based on the same word, which was continuously updated with each subsequent object they encountered. This prototypical representation might have helped infants recognize category members during the test phase and distinguish them from objects outside the category. Consequently, the prototypical information might have helped infants distinguish between objects within the category and objects outside the category during the test phase.

It is also possible that the infants formed a category because the consistent word highlighted concrete examples of objects belonging to that category, rather than on the basis of a prototype. In this case, infants would have stored individual examples of objects in a category in memory when they heard the same word for each object and used these examples to identify the novel object from the known category at test. More specifically, infants identify a novel object as a member of a category by comparing its similarity to memory representations of all previous examples from that category that were associated with the same name. For example, if infants have learned the name of several different types of stuffed daleks, they may store individual examples of each type of dalek (e.g., the red, purple, or blue dalek) and use these individual examples to form a "dalek" category that they can use to include the novel dalek from the familiar dalek category they see at test.

Although it is not clear from the data which strategy the infants actually used to form an object category based on a consistent word cue, the above strategies highlight that infants may interpret a consistent word as an additional feature in the visual scene that correlates with the visual similarities of objects belonging to the same category (Plunkett et al., 2008). From the feature-based perspective, then, a consistent word is in some sense comparable to consistent visual features, and thus can be treated similarly to other perceptual features of the object to distinguish between those that have the same name and those that do not. In this sense, the word can function in category formation as a feature that is not necessarily associated with the individual object or with all objects that belong to the same perceptual category.

The use of prototype and exemplar strategies in the formation of categories might have been more difficult in the presence of arbitrary actions performed on objects, because the actions might have distorted the visual perception of the objects and made it difficult for infants to recognize visual similarities among the objects that were moved the same (see also the Study II *Implications* section as part of this discussion for further arguments as to why actions may not elicit similar category effects to words). Such visual distortion of the moved object could have hindered the encoding of perceptual features of the objects, which in turn may have affected category formation. Actions, hence, induce perceptual variability to the visual scene which potentially account for visual object features, i.e., the the same action likely introduced comparable perceptual variations among the object presentations (i.e., all category objects were moved the same). As a result, infants in the action condition may have focused on exploring individual features of moved objects, which makes it likely that action-object presentations are processed in a bottom-up fashion in this context.

If both words and actions are considered as features of the objects, then both should have comparable effects on category formation. However, our data showed that only words, but not actions, facilitated the formation of a basic category at 1 year of age. There may hence be differences in the processes underlying category formation based on word-object versus action-object presentations. Such differences may arise due to additional cognitive factors contributing to the categorization process based on a word, but not based on an action cue. Words, then, likely play a role that goes beyond the function of object features as cues to category membership.

Category learning based on induction: A top-down approach

In Chapter 3, the data suggest that words may play a role in promoting infants' ability to recognize commonalities between objects beyond their visual attributes. Several studies suggest that consistently naming objects with a particular word can draw infants' attention to commonalities between objects with the same name (Althaus & Mareschal, 2014; Booth & Waxman, 2002; Fulkerson & Haaf, 2003;

Fulkerson & Waxman, 2007; Waxman & Markow, 1995; Waxman & Braun, 2005). This consistent labeling may highlight common perceptual features of objects with the same name, which in turn may lead to the formation of shared categories. As a result, infants' expectations of what they will find in a particular category based on prior experience may be shaped by the words they associate with those objects (Waxman & Markow, 1995; Waxman & Braun, 2005). That is, infants detect statistical regularities in the speech stream (e.g., Saffran, 2003) and are sensitive the regularities of perceptual features of objects (e.g., Havy & Waxman, 2016; Waxman & Markow, 1995). This sensitivity to the can help infants recognize commonalities between them and form shared categories and shape infants' expectations of what they will find in a particular category based on prior experience. Words, then, can manipulate infants' attention in the visual scene, drawing their focus to systematic visual structures like the similarities of objects.

Lupyan and his colleagues have taken the notion that words can enhance object perception a step further by proposing that words can actually influence how we perceive and categorize novel objects. They propose that word-object pairs can produce statistical regularities that cause objects to be perceived as more salient, leading to changes in perceptual space (e.g., Lupyan, 2012; Lupyan & Bergen, 2016; Lupyan & Thompson-Schill, 2012). For the infants who participated in the categorization study presented in Chapter 3, this could mean that the consistent labeling of each object in the study may have altered the infants' perception of the entire visual scene. The presence of the consistent word may have highlighted the object in the visual scene and made it visually salient and noticeable, potentially distorting the overall perception of the entire scene. In other words, the word-object pairs may have led the infants to see the scene differently by altering their attentional focus and object salience and enabled processes underlying category formation. While actions can also alter the visual scene by physically moving objects, it is less clear whether they can affect the perception of objects in the same way as suggested for words. Because the data reported in Chapter 3 showed that words, but not actions, have an effect on the formation of object categories in infants, it may be that the mechanisms by which words influence category formation are different from those underlying action-object relations in this context.

The above theories provide some explanations for why words, but not actions, may have a greater influence on the formation of object categories, as reported in Chapter 3. However, it remains puzzling that actions, which are similarly associated with objects, do not seem to have the same effect. It is possible that word processing and category formation are intertwined, so that these processes may reinforce each other. Indeed, there is some evidence to support this idea, as research by Pomiechowska & Gliga (2019) suggests that category knowledge can facilitate the learning of word-object associations. In their study, 12-month-old infants were shown pictures of objects from two different categories (staplers and coffee makers) in either blocked order (all objects from one category after all objects from the other category) or random order (objects from both categories were shown in random order), followed by naming one object per category while only the named object was displayed on the screen. At test, infants were presented with a stapler and a coffee maker side by side while they heard the name of one of the objects. They found that infants who saw objects of one category in a blocked scenario during familiarization were able to extend the name of the object to another member of that category (i.e., another stapler or coffee maker), whereas infants who saw random objects of the category during familiarization did not. This suggests that knowledge of categories may facilitate the learning of object names and that word learning and category learning may be interdependent and mutually reinforcing.

However, it is important to note that the above study did not examine whether the infants actually learned the word, which is a critical aspect of word extension at this developmental stage. Additionally, it remains unclear whether presenting objects within a category in a blocked manner is more beneficial for learning than randomizing the presentation. To address these issues and further explore the relationship between word learning and category formation, I conducted a study with 12-month-olds that is included as a brief report in Chapter 4 in this thesis[2]. The study includes both a word-learning task and a category formation task, with half of the infants completing the word task first and the other half completing the category task first. The first task was manipulated to allow for successful learning or not, enabling the measurement of the impact of successful learning in one task on the subsequent task. If word learning and category formation are interdependent, then infants who learn a word for an object

[2] The results of the study will be published in the context of this PhD shortly after thesis submission. Due to the COVID-19 pandemic, data collection for the study was interrupted and subsequently moved online. However, the results obtained from this online data collection were not included in the thesis, as there were significant differences in the infants' looking behavior when tested on a laptop screen at home compared to the results obtained in the laboratory, as reported in Chapter 3. Therefore, only eye-tracking data collected in the laboratory will be used for the analysis, and the analysis will be conducted once data collection is complete, as per the pre-registration for this study on the Open Science Framework. Currently, data collection for the study in the laboratory is 80% complete with a sample size of 64 out of 80 participants.

should be more likely to form a subsequent object category than if they do not learn the word for that object. If infants form an object category, they may also be better able to learn words associated with another member of that category. However, if these processes occur independently, infants should be able to form a subsequent object category whether or not they have learned the word for that object.

If the process of word learning and category formation are mutually reinforcing, as suggested by the literature (e.g., see Bothe et al., subm., and Fulkerson & Waxman, 2007, for evidence that words promote object categorization and Pomiechowska & Gliga, 2019, for how category knowledge may impact word learning), then it is likely that word effects in category learning are related to additional cognitive resources including attention allocation and memory. Thus, the integration of multiple physical objects into a unified mental representation based on a word cue may not only involve the exploration of visual features in a bottom-up manner, but also additional cognitive processes that enable infants to go beyond the surface-level features of objects. This suggests that category formation based on a consistent word is likely to involve both bottom-up and top-down cognitive processes, including the use of words as object category information as well as higher cognitive resources such as attention and memory. Such cognitive factors may play a unifying role in the process of word and category learning and may even give words a head start in object categorization. In the next section, I'll elaborate on what these cognitive factors could look like and how these factors connect word processing and category formation.

Word processing can stimulate and/or depend on cognitive control

The formation of categories based on consistent word cues probably involves both bottom-up and top-down cognitive processes, whereas the formation of categories based on consistent action cues may rely more heavily on bottom-up information. The idea I want to convey here is that because processes involved in word and category learning are closely interrelated and influence each other, words may have access to cognitive control that promote the formation of object categories. Cognitive control is a mental process which involved in executive control and decision making processes (Domenech & Koechlin, 2015) and learning (Koechlin, 2016). Such a link between word effects in category formation and cognitive control does not exclude words as features of categories, but it may be that additional levels of cognitive control provide words with a head start in stripping object similarities into a unified representation of the underlying category.

Enhanced cognitive control in category formation based on word cues may also explain why we found differences between words and actions in category formation (see Chapter 3) but not in association with objects (see Chapter 2), because additional levels of cognitive control are involved beyond the basic mechanisms of association and reinforcement learning (Rouault & Koechlin, 2018). In the context of the category formation data presented in Chapter 3, this may suggest that the consistent word cue triggered a higher level of cognitive control during the habituation phase in which infants were introduced to object members from a single category when accompanied by a word cue, but not when accompanied by an action. Cognitive control, then, may facilitate a number of cognitive processes, such as increased attentional allocation to the named objects, rapid identification of the word as an object feature, and improved ability to recognize object similarities.

By tapping into these complex cognitive mechanisms, words may provide an edge in early cognitive development. Indeed, research has shown that there is a strong relationship between language skills and cognitive control, while language development contributes to or depends on cognitive control. For example, studies have shown that language development predicts later cognitive development, which is reflected in educational attainment (McMurray, 2007). In addition, visual categorization of objects has been shown to be flexible and related to other cognitive domains. For example, infants with larger receptive vocabularies tend to categorize objects by both material and shape, whereas infants with smaller receptive vocabularies tend to categorize primarily by shape (Ellis & Oakes, 2006). Such findings suggest that language skills may allow for more detailed exploration of the visual input and strengthen the link between word and category learning. Exposing infants to multiple languages can enhance cognitive control skills, according to (Kovács & Mehler, 2009). Recent evidence by Gunnerud et al. (2020) suggests that while bilingualism may offer cognitive advantages, the relationship is not always consistent across studies. Nonetheless, exposure to a variety of phonological input from different languages and frequent switching between different vocabularies may promote cognitive advances in early development. Overall, the above findings suggest that the relationship between words and cognitive control is complex and multifaceted, and that language development may play a critical role in shaping cognitive processes in early infancy.

In summary, there is an intrinsic connection between words and cognitive control that could potentially explain some of the differences we observe between words and actions in the formation of a category as reported in Chapter 3. In what follows, I will address the overarching question of whether words are special in early development by summarizing the implications of the results presented in this thesis for the distinction between words and actions.

Outlook

Actions and words in early development

The research in this thesis highlights two distinct learning processes - associative learning and categorization, although the order in which they occur may not necessarily be fixed, as category formation may occur independently of associative processes. The first finding suggests that arbitrary actions and words have a similar impact on basic associative learning during early development, potentially reflecting shared mechanisms. The second finding indicated some evidence that only words, but not arbitrary actions, facilitated the formation of a basic object category. These results suggest that different processes may be at play - one for associative learning and another for generalizing within a category and distinguishing these objects within a category against objects outside that category. It is possible that category formation requires more sophisticated processing abilities than simple associative learning. If associative processes are indeed less complex and potentially precede category learning, infants may find it easier to associate an action with an object than to identify the action as a cue to an object category. Actions are a frequent part of infants' input and allow infants to communicate with others before they speak, which precedes and sets the stage for verbal communication. This communication trajectory is crucial since nonverbal cues, such as actions and facial expressions, convey emotions, attitudes, and intentions, and provide context for verbal communication which allows for a more effective exchange of information (e.g., Iverson & Goldin-Meadow, 2005).

However, advanced cognitive tasks may favor the use of words over actions, providing verbal communication (e.g., object naming) with an advantage in structuring and conceptualizing visual information when nonverbal cues are insufficient to process the entirety of the visual information. For instance, one difference between category learning and word-object and action-object mappings in early development is the need for longer attentional spans and memory capacities during category formation. To form categories, infants must attend to multiple objects and detect commonalities among them, which requires a greater cognitive effort than attending to a single object and its corresponding cue in simpler mappings. If words are closely linked with cognitive functions such as cognitive control, which enable access to additional attention and memory resources, then words may be more conducive to promoting category formation. In contrast, actions may deplete these resources during category formation because attending to a moving object may impose greater cognitive demands on object encoding compared to when the object is presented still.

These findings could have practical implications for caregivers who seek to enhance their infants' learning and development. Based on the results presented in Chapter 3, caregivers can be encouraged to name objects that infants may see often and in different contexts to promote their category learning. In addition, findings from the associative learning study in Chapter 2 showed that caregivers can use their hands to point to or touch the objects that are being named in a synchronous manner, which has been shown to facilitate infants' learning of novel associations. By doing so, caregivers can potentially enhance their infants' associative learning and promote their early cognitive development.

Words are special in early development because they uniquely promote category formation

Taken together, this thesis contributes significantly to the literature on the role of words in early development. The results suggest that words play a unique role in early development when it comes to the formation of categories. This is underscored by the fact that actions comparable to words in terms of familiarity and arbitrariness of word-object relations did not show a similar pattern of effects. Specifically, two main results have emerged from this research. Firstly, infants either associate both, words with objects and actions with objects, or they did not associate either of the cues with objects.

This latter was observed when word-action-object triads were introduced sequentially (i.e., one after the other) or when infants were young. Secondly, words appear to have an advantage over other cues, not only within the auditory modality (like sounds), but also compared to arbitrary actions across modalities, during the process of object categorization. This unique advantage was not previously identified in the literature, as previous comparisons of words were made with functions (Booth & Waxman, 2002) or self-initiated movements (Sučević et al., 2021). The present thesis demonstrates that when word cues and action cues have an equal arbitrary relationship with the object, there is evidence to suggest that words alone may simplify the categorization process. It is therefore plausible that words hold a unique status in the activation of cognitive resources that facilitate effective similarity detection, generalized perception of object features, and ultimately the categorization of physical objects.

Conclusion

This thesis aimed to investigate the distinct role of words in early development, given their ability to promote the formation of physical object categories, which is not found in other auditory cues like sounds. Because words are highly familiar to infants, the study compared the effect of words to equally familiar cues across modalities, namely actions. These action cues were matched to the arbitrariness of the word-object relationship, meaning that the action did not convey a function or intention. Previous research has shown that infants do not learn words for objects when actions accompany words in laboratory settings when these cues were presented synchronously with the objects. Since word-action-object presentations are typically synchronized in infants' natural interactions with caregivers, we reevaluate infants' associative learning success again under two temporal conditions. In the first task, infants' recognition of objects based on word and action cues was tested after exposure to the word-action-object triads, presented either synchronously or sequentially. Results from the associative learning study revealed that 2-year-olds learned both words and actions for objects when introduced synchronously, but not sequentially, as determined by looking time measures. However, these effects were not confirmed by pupil measures in a subsequent violation of expectation task, suggesting that learning effects were transient and did not hold until the second test phase. In contrast, younger infants aged 1 year did not show learning for either cue, possibly due to the additional costs of multisensory processing in the laboratory environment that disrupted their learning of the associations.

In contrast to the associative learning study, where 1-year-olds either learned both words and actions for objects or none of the cues, our results indicated that word cues promoted category formation while action cues did not. These findings suggest that there are distinct roles for words and actions in promoting the individuation of objects through associative learning and the categorization of objects based on feature generalization. Specifically, while both single word and action cues led infants to habituate to a single object category, only words reliably provoked novelty preference at test, reflecting learning of the category. Infants exposed to both words and actions with in-category objects did not exhibit evidence of habituation or categorization, possibly due to the greater demands of multisensory processing in a category study where multiple objects are presented over a prolonged period. Overall, these results indicate that words play a significant role in the relational learning and grouping of objects, allowing commonalities to be stripped into a single mental representation of physical entities.

The current thesis, thus, provides novel insights into the role of words in early development and highlights the unique contribution of words to infant cognitive development. Specifically, the findings suggest that the way that words are associated with objects is similarly the way that equally familiar and arbitrary cues like actions, are associated with words, suggesting that there are similarities underlying word-object and action-object learning. However, we found word effects in category formation which were not reproduced by the equally familiar and arbitrary actions, suggesting that words provide a unique advantage in promoting learning during early development. This finding is particularly striking because such an advantage of words in category formation persists even when action cues are matched in arbitrariness. This suggests that words facilitate learning and the development of more advanced conceptual abilities in a manner that cannot be replicated by similar cues across modalities. Overall, this thesis highlights the crucial role of words in early development and provides new avenues for future research into the mechanisms underlying early learning and cognition.

References

Ackermann, L., Hepach, R., & Mani, N. (2019). Children learn words easier when they are interested in the category to which the word belongs. *Developmental Science*, desc.12915. https://doi.org/10.1111/desc.12915

Ackermann, L., Hepach, R., & Mani, N. (2020). Children learn words easier when they are interested in the category to which the word belongs. *Developmental Science*, *23*(3). https://doi.org/10.1111/desc.12915

Allen, R. J., Baddeley, A. D., & Hitch, G. J. (2006). Is the binding of visual features in working memory resource-demanding? *Journal of Experimental Psychology: General*, *135*(2), 298–313. https://doi.org/10.1037/0096-3445.135.2.298

Althaus, N., Gliozzi, V., Mayor, J., & Plunkett, K. (2020). Infant categorization as a dynamic process linked to memory. *Royal Society Open Science*, *7*(10), 200328. https://doi.org/10.1098/rsos.200328

Althaus, N., & Mareschal, D. (2014). Labels Direct Infants' Attention to Commonalities during Novel Category Learning. *PLoS ONE*, *9*(7), e99670. https://doi.org/10.1371/journal.pone.0099670

Althaus, N., & Plunkett, K. (2015). Timing matters: The impact of label synchrony on infant categorisation. *Cognition*, *139*, 1–9. https://doi.org/10.1016/j.cognition.2015.02.004

Althaus, N., & Westermann, G. (2016). Labels constructively shape object categories in 10-month-old infants. *Journal of Experimental Child Psychology*, *151*, 5–17. https://doi.org/10.1016/j.jecp.2015.11.013

Arbib, M. A. (2006). *Action to language via the mirror neuron system*. Cambridge University Press. http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&AN=171856

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

Axelsson, E. L., Churchley, K., & Horst, J. S. (2012). The Right Thing at the Right Time: Why Ostensive Naming Facilitates Word Learning. *Frontiers in Psychology*, *3*(88). https://doi.org/10.3389/fpsyg.2012.00088

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

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. https://doi.org/10.1111/1467-8624.00495

Bahrick, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental Psychology*, *36*(2), 190–201. https://doi.org/10.1037/0012-1649.36.2.190

Bahrick, L. E., & Lickliter, R. (2004). Infants' perception of rhythm and tempo in unimodal and multimodal stimulation: A developmental test of the intersensory redundancy hypothesis. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(2), 137–147. https://doi.org/10.3758/CABN.4.2.137

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. https://doi.org/10.1111/j.0963-7214.2004.00283.x

Bahrick, L. E., & Newell, L. C. (2008). Infant discrimination of faces in naturalistic events: Actions are more salient than faces. *Developmental Psychology*, *44*(4), 983–996. https://doi.org/10.1037/0012-1649.44.4.983

Baillargeon, R. (2004). Can 12 large clowns fit in a Mini Cooper? Or when are beliefs and reasoning explicit and conscious? *Developmental Science*, *7*(4), 422–424. https://doi.org/10.1111/j.1467-7687.2004.00361.x

Baillargeon, R., Spelke, E. S., & Wasserman, S. (1985). Object permanence in five-month-old infants. *Cognition*, *20*(3), 191–208. https://doi.org/10.1016/0010-0277(85)90008-3

Balaban, M. T., & Waxman, S. R. (1997). Do Words Facilitate Object Categorization in 9-Month-Old Infants? *Journal of Experimental Child Psychology*, *64*(1), 3–26. https://doi.org/10.1006/jecp.1996.2332

Ballem, K. D., & Plunkett, K. (2005). Phonological specificity in children at 1;2. *Journal of Child Language*, *32*(1), 159–173. https://doi.org/10.1017/S0305000904006567

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, *4*(328). https://doi.org/10.3389/fpsyg.2013.00328

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bates, E., Bretherton, I., Snyder, L., Shore, C., & Volterra, V. (1980). VOCAL AND GESTURAL SYMBOLS AT 13 MONTHS. *Merrill-Palmer Quarterly of Behavior and Development*, *26*(4), 407–423.

Bates, E., Dale, P. S., & Thal, D. (1995). Individual Differences and their Implications for Theories of Language Development. In P. Fletcher & B. MacWhinney (Eds.), *The Handbook of Child Language* (pp. 95–151). Blackwell Publishing Ltd. https://doi.org/10.1111/b.9780631203124.1996.00005.x

Beatty, J. (1982). Task-Evoked Pupillary Responses, Processing Load, and the Structure of Processing Resources. *Psychological Bulletin*, *91*(2), 276–292.

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. https://doi.org/10.1073/pnas.1113380109

Bloom, L. (1993). *The transition from infancy to language: Acquiring the power of expression. Cambridge, U.K.: Cambridge University Press.* Cambridge, U.K.: Cambridge University Press.

Bloom, P. (2000). *How children learn the meanings of words*. MIT Press.

Bolker, B. (2008). *Ecological Models and Data in R*. Princeton University Press.

Bomba, P. C., & Siqueland, E. R. (1983). The nature and structure of infant form categories. *Journal of Experimental Child Psychology*, *35*(2), 294–328. https://doi.org/10.1016/0022-0965(83)90085-1

Booth, A. E., & Waxman, S. (2002). Object names and object functions serve as cues to categories for infants. *Developmental Psychology*, *38*(6), 948–957. https://doi.org/10.1037/0012-1649.38.6.948

Booth, A. E., & Waxman, S. R. (2009). A Horse of a Different Color: Specifying With Precision Infants' Mappings of Novel Nouns and Adjectives. *Child Development*, *80*(1), 15–22. https://doi.org/10.1111/j.1467-8624.2008.01242.x

Booth, A., & Waxman, S. (2002). Word learning is "smart": Evidence that conceptual information affects preschoolers' extension of novel words. *Cognition*, *84*(1), B11–B22. https://doi.org/10.1016/S0010-0277(02)00015-X

Borgström, K., Torkildsen, J. von K., & Lindgren, M. (2015). Substantial gains in word learning ability between 20 and 24 months: A longitudinal ERP study. *Brain and Language*, *149*, 33–45. https://doi.org/10.1016/j.bandl.2015.07.002

Borovsky, A., & Elman, J. (2006). Language input and semantic categories: A relation between cognition and early word learning. *Journal of Child Language*, *33*(4), 759–790. https://doi.org/10.1017/S0305000906007574

Bothe, R., Eiteljoerge, S. F. V., Trouillet, L., Elsner, B., & Mani, N. (under review). Better in sync: Temporal dynamics explain multisensory word-action-object learning in early development. *Infancy*.

Bothe, R., Trouillet, L., Elsner, B., & Mani, N. (submitted). Can words do more than arbitrary actions in early early category formation in infancy? *Developmental Science*.

Brand, R. J., Baldwin, D. A., & Ashburn, L. A. (2002). Evidence for 'motionese': Modifications in mothers' infant-directed action: Infant-directed action. *Developmental Science*, *5*(1), 72–83. https://doi.org/10.1111/1467-7687.00211

Brand, R. J., Shallcross, W. L., Sabatos, M. G., & Massie, K. P. (2007). Fine-Grained Analysis of Motionese: Eye Gaze, Object Exchanges, and Action Units in Infant-Versus Adult-Directed Action. *Infancy*, *11*(2), 203–214. https://doi.org/10.1111/j.1532-7078.2007.tb00223.x

Brand, R., & Shallcross, W. (2008). Infants prefer motionese to adult-directed action. *Developmental Science*, *11*(6), 853–861. https://doi.org/DOI: 10.1111/j.1467-7687.2008.00734.x

Brandone, A. C., Horwitz, S. R., Aslin, R. N., & Wellman, H. M. (2014). Infants' goal anticipation during failed and successful reaching actions. *Developmental Science*, *17*(1), 23–34. https://doi.org/10.1111/desc.12095

Brandone, A. C., & Wellman, H. M. (2009). You Can't Always Get What You Want. *Psychological Science*, *20*(1), 85–91.

Broadbent, H. J., Osborne, T., Rea, M., Peng, A., Mareschal, D., & Kirkham, N. Z. (2018). Incidental category learning and cognitive load in a multisensory environment across childhood. *Developmental Psychology*, *54*(6), 1020–1028. https://doi.org/10.1037/dev0000472

Broemeling, L. D., & Good, I. J. (2007). *Bayesian sequential analysis*. John Wiley & Sons.

Brooks, M., E., Kristensen, K., Benthem, K., J. ,van, Magnusson, A., Berg, C., W., Nielsen, A., Skaug, H., J., Mächler, M., & Bolker, B., M. (2017). GlmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal*, *9*(2), 378–400. https://doi.org/10.32614/RJ-2017-066

Brown, R. (1973). *A First Language: The Early Stages.* Cambridge, MA: Harvard Universirty Press.

Byers-Heinlein, K. (2015). Methods for studying infant bilingualism. In J. W. Schwieter (Ed.), *The Cambridge Handbook of Bilingual Processing* (pp. 133–154). Cambridge University Press. https://doi.org/10.1017/CBO9781107447257.005

Byers-Heinlein, K., Bergmann, C., & Savalei, V. (2021). *Six solutions for more reliable infant research* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/u37fy

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

Capirci, A., Contaldo, A., Caselli, M. C., & Volterra, V. (2005). From action to language through gesture: A longitudinal perspective. *Gesture*, *5*, 77–155.

Caselli, M. C., Rinaldi, P., Stefanini, S., & Volterra, V. (2012). Early Action and Gesture "Vocabulary" and Its Relation With Word Comprehension and Production: **Early Action and Gesture "Vocabulary."** *Child Development*, *83*(2), 526–542. https://doi.org/10.1111/j.1467-8624.2011.01727.x

Cattani, A., Floccia, C., Kidd, E., Pettenati, P., Onofrio, D., & Volterra, V. (2019). Gestures and Words in Naming: Evidence From Crosslinguistic and Crosscultural Comparison. *Language Learning*, *69*(3), 709–746. https://doi.org/10.1111/lang.12346

Chatterjee, S., & Simonoff, J. S. (2013). *Handbook of regression analysis*. Wiley.

Cheng, C., Kaldy, Z., & Blaser, E. (2019). Focused attention predicts visual working memory performance in 13-month-old infants: A pupillometric study. *Developmental Cognitive Neuroscience*, *36*, 100616. https://doi.org/10.1016/j.dcn.2019.100616

Cheung, R. W., Hartley, C., & Monaghan, P. (2021). Caregivers use gesture contingently to support word learning. *Developmental Science*, *24*(4). https://doi.org/10.1111/desc.13098

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

Christie, J., & Klein, R. (1995). Familiarity and attention: Does what we know affect what we notice? *Memory & Cognition*, *23*(5), 547–550. https://doi.org/10.3758/BF03197256

Clark, E. (2016). *First Language Acquisition*. 3rd ed. Cambridge, UK: Cambridge University Press.

Cohen, L. B. (2004). Uses and misuses of habituation and related preference paradigms. *Infant and Child Development*, *13*(4), 349–352. https://doi.org/10.1002/icd.355

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

DeLoache, J. S. (2004). Becoming symbol-minded. *Trends in Cognitive Sciences*, *8*(2), 66–70. https://doi.org/10.1016/j.tics.2003.12.004

Deng, W. (Sophia), & Sloutsky, V. (2011). *Carrot-Eaters and Moving Heads: Salient Features Provide Greater Support for Inductive Inference than Category Labels*.

Dobson, A. J. (2002). *An introduction to generalized linear models* (2nd ed). Chapman & Hall/CRC.

Domenech, P., & Koechlin, E. (2015). Executive control and decision-making in the prefrontal cortex. *Current Opinion in Behavioral Sciences*, *1*, 101–106. https://doi.org/10.1016/j.cobeha.2014.10.007

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. https://doi.org/10.1016/j.jecp.2016.01.003

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

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. https://doi.org/10.1371/journal.pone.0220317

Ellis, A. E., & Oakes, L. M. (2006). Infants flexibly use different dimensions to categorize objects. *Developmental Psychology*, *42*(6), 1000–1011. https://doi.org/10.1037/0012-1649.42.6.1000

Elsner, B., & Aschersleben, G. (2004). Do I get what you get? Learning about effects of self-performed and observed actions in infants. *Consciousness and Cognition*, *12*, 732–751. https://doi.org/10.1016/S1053-8100(03)00073-4

Emberson, L. L., Misyak, J. B., Schwade, J. A., Christiansen, M. H., & Goldstein, M. H. (2019). Comparing statistical learning across perceptual modalities in infancy: An investigation of underlying learning mechanism(s). *Developmental Science*, *22*(6). https://doi.org/10.1111/desc.12847

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

Feldman, R. (2012). Parent-infant synchrony: A biobehavioral model of mutual influences in the formation of affiliative bonds. *Soc. Res. Child Dev.*, *77*, 42–51.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C. B., & Stiles, J. (1994). Variability in Early Communicative Development. *Monographs of the Society for Research in Child Development*, *59*(5). https://doi.org/10.2307/1166093

Fenson, L., Marchman, V., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). MacArthur–Bates CDI (Communicative Development Inventory) words and sentences. *Baltimore: Brookes Publishing*.

Ferguson, B., & Waxman, S. (2017). Linking language and categorization in infancy. *Journal of Child Language*, *44*(3), 527–552. https://doi.org/10.1017/S0305000916000568

Ferry, A. L., Hespos, S. J., & Waxman, S. R. (2010). Categorization in 3- and 4-Month-Old Infants: An Advantage of Words Over Tones. *Child Development*, *81*(2), 472–479. https://doi.org/10.1111/j.1467-8624.2009.01408.x

Field, A. (2005). Regression. *Discovering Statistics Using SPSS*, *2*, 143–217.

Forstmeier, W., & Schielzeth, H. (2011). Cryptic multiple hypotheses testing in linear models: Overestimated effect sizes and the winner's curse. *Behavioral Ecology and Sociobiology*, *65*(1), 47–55. https://doi.org/10.1007/s00265-010-1038-5

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. (2016). Wordbank: An open repository for developmental vocabulary data*. *Journal of Child Language*, 1–18.

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. (2020). *Variability and Consistency in Early Language Learning The Wordbank Project*.

Fulkerson, A. L., & Haaf, R. A. (2003). The Influence of Labels, Non-Labeling Sounds, and Source of Auditory Input on 9- and 15-Month-Olds' Object Categorization. *Infancy*, *4*(3), 349–369. https://doi.org/10.1207/S15327078IN0403_03

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. https://doi.org/10.1016/j.cognition.2006.09.005

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis*. CRC press.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Ghahramani, Z. (2015). Probabilistic machine learning and artificial intelligence. *Nature*, *521*(7553), 452–459. https://doi.org/10.1038/nature14541

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.

Gogate, L. (2018). *Multisensoric perception and communication (brain, behavior, environmental interaction and development in the early years)*. London: Routledge.

Gogate, L. (2020). Maternal object naming is less adapted to preterm infants’ than to term infants’ word mapping. *Journal of Child Psychology and Psychiatry*, *61*(4), 447–458. https://doi.org/10.1111/jcpp.13128

Gogate, L., & Hollich, G. (Eds.). (2013). *Theoretical and Computational Models of Word Learning: Trends in Psychology and Artificial Intelligence*. IGI Global. https://doi.org/10.4018/978-1-4666-2973-8

Gogate, L. J. (2010). Learning of syllable–object relations by preverbal infants: The role of temporal synchrony and syllable distinctiveness. *Journal of Experimental Child Psychology*, *105*(3), 178–197. https://doi.org/10.1016/j.jecp.2009.10.007

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. https://doi.org/10.1006/jecp.1998.2438

Gogate, L. J., & Bahrick, L. E. (2001). Intersensory Redundancy and 7-Month-Old Infants’ Memory for Arbitrary Syllable-Object Relations. *Infancy*, *2*(2), 219–231. https://doi.org/10.1207/S15327078IN0202_7

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. https://doi.org/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. https://doi.org/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. https://doi.org/10.1037/a0019049

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. https://doi.org/10.1111/1467-7687.00143

Gogate, L., Maganti, M., & Perenyi, A. (2014b). Preterm and Term Infants' Perception of Temporally Coordinated Syllable–Object Pairings: Implications for Lexical Development. *Journal of Speech, Language, and Hearing Research*, *57*(1), 187–198. https://doi.org/10.1044/1092-4388(2013/12-0403)

Goldin-Meadow, S., Alibali, M. W., & Church, R. B. (1993). Transitions in concept acquisition: Using the hand to read the mind. *Psychological Review*, *100*(2), 279–297. https://doi.org/10.1037/0033-295X.100.2.279

Goldin-Meadow, S., & Brentari, D. (2017). Gesture, sign, and language: The coming of age of sign language and gesture studies. *Behavioral and Brain Sciences*, *40*, e46. https://doi.org/10.1017/S0140525X15001247

Goldstein, M., Elmlinger, S., & Schwade, J. (2020, July 6). *Relative contributions of infant-directed speech and motion when learning new words* [Conference presentation]. vICIS 2020 Congress. https://infantstudies.org/congress-2020/

Golinkoff, R. M., Can, D. D., Soderstrom, M., & Hirsh-Pasek, K. (2015). (Baby)Talk to Me: The Social Context of Infant-Directed Speech and Its Effects on Early Language Acquisition. *Current Directions in Psychological Science*, *24*(5), 339–344. https://doi.org/10.1177/0963721415595345

Golinkoff, R. M., Ma, W., Song, L., & Hirsh-Pasek, K. (2013). Twenty-Five Years Using the Intermodal Preferential Looking Paradigm to Study Language Acquisition: What Have We Learned? *Perspectives on Psychological Science*, *8*(3), 316–339. https://doi.org/10.1177/1745691613484936

Gredebäck, G., Johnson, S., & Hofsten, C. (2009). Eye tracking in infancy research. *Developmental Neuropsychology*, *35*(1), 1–19.

Green, P., & MacLeod, C. J. (2016). SimR: an R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Grimm, H., & Doil, H. (2000). *Elternfragebögen für die Früherkennung von Risikokindern: ELFRA.* Hogrefe, Verlag für Psychologie.

Gunnerud, H. L., ten Braak, D., Reikerås, E. K. L., Donolato, E., & Melby-Lervåg, M. (2020). Is bilingualism related to a cognitive advantage in children? A systematic review and meta-analysis. *Psychological Bulletin*, *146*(12), 1059–1083. https://doi.org/10.1037/bul0000301

Hall, D. G., & Waxman, S. (1993). Assumptions about Word Meaning: Individuation and Basic-Level Kinds. *Child Development*, *64*, 1550–1570.

Havy, M., & Waxman, S. R. (2016). Naming influences 9-month-olds' identification of discrete categories along a perceptual continuum. *Cognition*, *156*, 41–51. https://doi.org/10.1016/j.cognition.2016.07.011

Hepach, R., Vaish, A., & Tomasello, M. (2012). Young Children Are Intrinsically Motivated to See Others Helped. *Psychological Science*, *23*(9), 967–972. https://doi.org/10.1177/0956797612440571

Hepach, R., & Westermann, G. (2016). Pupillometry in Infancy Research. *Journal of Cognition and Development*, *17*(3), 359–377. https://doi.org/10.1080/15248372.2015.1135801

Herschkowitz, N. N. (1988). Brain development in the fetus, neonate and infant. *Biology of the Neonate*, *54*(1), 1–19.

Hollich G. J. Golinkoff R. M. Hirsh-Pasek K. & Society for Research in Child Development. (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Blackwell.*

Horowitz, A. (2009). Attention to attention in domestic dog (Canis familiaris) dyadic play. *Animal Cognition*, *12*(1), 107–118. https://doi.org/10.1007/s10071-008-0175-y

Houston-Price, C., Mather, E., & Sakkalou, E. (2007). Discrepancy between parental reports of infants' receptive vocabulary and infants' behaviour in a preferential looking task. *Journal of Child Language*, *34*(4), 701–724. https://doi.org/10.1017/S0305000907008124

Hu, J. F. (2008). *The impact of labelling on categorisation processes in infancy*. PhD thesis, Department of Experimental Psychology, Oxford University.

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

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in Infancy Research*, *5*, 69–95.

Iverson, J. M., Capirci, O., & Caselli, M. C. (1994). From communication to language in two modalities. *Cognitive Development*, *9*(1), 23–43. https://doi.org/10.1016/0885-2014(94)90018-3

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture Paves the Way for Language Development. *Psychological Science*, *16*(5), 367–371. https://doi.org/10.1111/j.0956-7976.2005.01542.x

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

Kaminski, J., Call, J., & Fischer, J. (2004). Word Learning in a Domestic Dog: Evidence for "Fast Mapping." *Science*, *304*(5677), 1682–1683. https://doi.org/10.1126/science.1097859

Kanakogi, Y., & Itakura, S. (2011). Developmental correspondence between action prediction and motor ability in early infancy. *Nature Communications*, *2*(1). https://doi.org/10.1038/ncomms1342

Karatekin, C. (2004). Development of attentional allocation in the dual task paradigm. *International Journal of Psychophysiology*, *52*(1), 7–21. https://doi.org/10.1016/j.ijpsycho.2003.12.002

Karatekin, C. (2007). Eye tracking studies of normative and atypical development. *Developmental Review*, *27*(3), 283–348. https://doi.org/10.1016/j.dr.2007.06.006

Karmiloff-Smith, A. (1997). Crucial differences between developmental cognitive neuroscience and adult neuropsychology. *Developmental Neuropsychology*, *13*(4), 513–524. https://doi.org/10.1080/87565649709540693

Kemler Nelson, D. G., Russell, R., Duke, N., & Jones, K. (2000). Two-Year-Olds Will Name Artifacts by Their Functions. *Child Development*, *71*(5), 1271–1288. https://doi.org/10.1111/1467-8624.00228

Koechlin, E. (2016). Prefrontal executive function and adaptive behavior in complex environments. *Current Opinion in Neurobiology*, *37*, 1–6. https://doi.org/10.1016/j.conb.2015.11.004

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. https://doi.org/10.1016/j.infbeh.2009.07.003

Kovács, Á. M., & Mehler, J. (2009). Cognitive gains in 7-month-old bilingual infants. *Proceedings of the National Academy of Sciences*, *106*(16), 6556–6560. https://doi.org/10.1073/pnas.0811323106

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

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of Neuronal Oscillations as a Mechanism of Attentional Selection. *Science*, *320*(5872), 110–113. https://doi.org/10.1126/science.1154735

Landau, B., Smith, L. B., & Jones, S. (1992). Syntactic context and the shape bias in children's and adults' lexical learning. *Journal of Memory and Language*, *31*(6), 807–825. https://doi.org/10.1016/0749-596X(92)90040-5

Lany, J., Aguero, A., & Thompson, A. (2020, July 6). *The temporal dynamics of labeling shape object recognition* [Conference presentation]. vICIS 2020 Congress. https://infantstudies.org/congress-2020/

Lany, J., Aguero, A., & Thompson, A. (2022). The temporal dynamics of labelling shape infant object recognition. *Infant Behavior and Development*, *67*, 101698. https://doi.org/10.1016/j.infbeh.2022.101698

Lany, J., Thompson, A., & Aguero, A. (2022). What's in a name, and when can a [beep] be the same? *Developmental Psychology*, *58*(2), 209–221. https://doi.org/10.1037/dev0001084

LaTourrette, A., & Waxman, S. R. (2019). A little labeling goes a long way: Semi-supervised learning in infancy. *Developmental Science*, *22*(1), e12736. https://doi.org/10.1111/desc.12736

Law, J., & Roy, P. (2008). Parental Report of Infant Language Skills: A Review of the Development and Application of the Communicative Development Inventories. *Child and Adolescent Mental Health*, *13*(4), 198–206. https://doi.org/10.1111/j.1475-3588.2008.00503.x

Lecoutre, B., & Poitevineau, J. (2011). *Sequential analysis and optimal design*. CRC Press.

Leong, V., Byrne, E., Clackson, K., Georgieva, S., Lam, S., & Wass, S. (2017). Speaker gaze increases information coupling between infant and adult brains. *Proceedings of the National Academy of Sciences*, *114*(50), 13290–13295. https://doi.org/10.1073/pnas.1702493114

Lewkowicz, D. (1994). Limitations on Infants' Response to Rate-Based Auditory-Visual Relations. *Developmental Psychology*, *30*, 880–892. https://doi.org/10.1037/0012-1649.30.6.880

Lewkowicz, D. J. (1988a). Sensory dominance in infants: I. Six-month-old infants' response to auditory-visual compounds. *Developmental Psychology*, *24*(2), 155-171.

Lewkowicz, D. J. (1988b). Sensory dominance in infants: II. Ten-month-old infants' response to auditory-visual compounds. *Developmental Psychology*, *24*(2), 172–182. https://doi.org/10.1037/0012-1649.24.2.172

Lewkowicz, D. J., & Ghazanfar, A. A. (2006). The decline of cross-species intersensory perception in human infants. *Proceedings of the National Academy of Sciences*, *103*(17), 6771–6774. https://doi.org/10.1073/pnas.0602027103

Lewkowicz, D. J., & Kraebel, K. S. (2004). The Value of Multisensory Redundancy in the Development of Intersensory Perception. *In G. A. Calvert, C. Spence, & B. E. Stein (Eds.), The Handbook of Multisensory Processes. MIT Press.*, 655–678.

Lidz, J., Waxman, S., & Freedman, J. (2003). What infants know about syntax but couldn't have learned: Experimental evidence for syntactic structure at 18 months. *Cognition*, *89*(3), 295–303. https://doi.org/10.1016/S0010-0277(03)00116-1

Lippe, S., Kovacevic, N., & Randal McIntosh, A. (2009). Differential maturation of brain signal complexity in the human auditory and visual system. *Frontiers in Human Neuroscience*, *3*. https://doi.org/10.3389/neuro.09.048.2009

Logan, J. A. R., Justice, L. M., Yumuş, M., & Chaparro-Moreno, L. J. (2019). When Children Are Not Read to at Home: The Million Word Gap. *Journal of Developmental & Behavioral Pediatrics*, *40*(5), 383–386. https://doi.org/10.1097/DBP.0000000000000657

Low, R., & Sweller, J. (2005). The Modality Principle in Multimedia Learning. In *The Cambridge handbook of multimedia learning.* (pp. 147–158). Cambridge University Press. https://doi.org/10.1017/CBO9780511816819.010

Luchkina, E., & Waxman, S. (2021). Acquiring verbal reference: The interplay of cognitive, linguistic, and general learning capacities. *Infant Behavior and Development*, *65*, 101624. https://doi.org/10.1016/j.infbeh.2021.101624

Lupyan, G. (2012). What Do Words Do? Toward a Theory of Language-Augmented Thought. In *Psychology of Learning and Motivation* (Vol. 57, pp. 255–297). Elsevier. https://doi.org/10.1016/B978-0-12-394293-7.00007-8

Lupyan, G., & Bergen, B. (2016). How Language Programs the Mind. *Topics in Cognitive Science*, *8*(2), 408–424. https://doi.org/10.1111/tops.12155

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not Just for Talking: Redundant Labels Facilitate Learning of Novel Categories. *Psychological Science*, *18*(12), 1077–1083. https://doi.org/10.1111/j.1467-9280.2007.02028.x

Lupyan, G., & Thompson-Schill, S. L. (2012). The evocative power of words: Activation of concepts by verbal and nonverbal means. *Journal of Experimental Psychology: General*, *141*(1), 170–186. https://doi.org/10.1037/a0024904

Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. *Proceedings of the National Academy of Sciences*, *110*(35), 14196–14201. https://doi.org/10.1073/pnas.1303312110

MacKenzie, H., Graham, S. A., & Curtin, S. (2011). Twelve-month-olds privilege words over other linguistic sounds in an associative. *Developmental Science*, *14*(2), 249–255.

Maganti, M., & Gogate, L. (2013). *Maternal synchronous gesture adaptations during object naming to term and preterm infants A longitudinal study* [Conference presentation]. vICIS 2020 Congress. https://infantstudies.org/congress-2020/

Mani, N., Schreiner, M. S., Brase, J., Köhler, K., Strassen, K., Postin, D., & Schultze, T. (2021). Sequential Bayes Factor designs in developmental research: Studies on early word learning. *Developmental Science 24*(4). https://doi.org/10.1111/desc.13097

Mareschal, D., & Quinn, P. C. (2001). Categorization in infancy. *Trends in Cognitive Sciences*, *5*(10), 443–450. https://doi.org/10.1016/S1364-6613(00)01752-6

Masataka, N. (1998). Perception of Motherese in Japanese Sign Language by 6-Month-Old Hearing Infants. *Developmental Psychology*, *34*(22), 241–246.

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. https://doi.org/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. https://doi.org/10.1016/j.jecp.2013.09.010

Mayor, J., & Mani, N. (2019). A short version of the MacArthur–Bates Communicative Development Inventories with high validity. *Behavior Research Methods*, *51*(5), 2248–2255. https://doi.org/10.3758/s13428-018-1146-0

McCullagh, P., & Nelder, J. (1989). *Generalized linear models*. Chapman and Hall.

McMurray, B. (2007). Defusing the Childhood Vocabulary Explosion. *Science*, *317*(5838), 631–631. https://doi.org/10.1126/science.1144073

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, *119*(4), 831–877. https://doi.org/10.1037/a0029872

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, *31*(5), 838–850. https://doi.org/10.1037/0012-1649.31.5.838

Mirman, D. (2016). *Growth curve analysis and visualization using R*. CRC press.

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. https://doi.org/10.1016/j.jml.2007.11.006

Misyak, J. B., & Christiansen, M. H. (2012). Statistical Learning and Language: An Individual Differences Study: Individual Differences in Statistical Learning. *Language Learning*, *62*(1), 302–331. https://doi.org/10.1111/j.1467-9922.2010.00626.x

Monroy, C., Gerson, S., & Hunnius, S. (2017). Infants' Motor Proficiency and Statistical Learning for Actions. *Frontiers in Psychology*, *8,* 2174. https://doi.org/10.3389/fpsyg.2017.02174

Nguyen, T., Zimmer, L., & Hoehl, S. (2022). *Your turn, my turn. Neural synchrony in mother-infant proto-conversation* [Preprint]. Neuroscience. https://doi.org/10.1101/2022.09.09.507239

Oakes, L. M., Plumert, J. M., Lansink, J. M., & Merryman, J. D. (1996). Evidence for task-dependent categorization in infancy. *Infant Behavior and Development*, *19*(4), 425–440. https://doi.org/10.1016/S0163-6383(96)90004-1

Oakes, L. M., & Ribar, R. J. (2005). A Comparison of Infants' Categorization in Paired and Successive Presentation Familiarization Tasks. *Infancy*, *7*(1), 85–98. https://doi.org/10.1207/s15327078in0701_7

Oakes, L. M., & Spalding, T. L. (1997). The role of exemplar distribution in infants' differentiation of categories. *Infant Behavior and Development*, *20*(4), 457–475.

Özçalışkan, Ş., & Goldin-Meadow, S. (2005). Gesture is at the cutting edge of early language development. *Cognition*, *96*(3), B101–B113. https://doi.org/10.1016/j.cognition.2005.01.001

Pätzold, W., & Liszkowski, U. (2019). Pupillometry reveals communication-induced object expectations in 12- but not 8-month-old infants. *Developmental Science, 22*(6). https://doi.org/10.1111/desc.12832

Pätzold, W., & Liszkowski, U. (2020). Pupillometric VoE paradigm reveals that 18- but not 10-month-olds spontaneously represent occluded objects (but not empty sets). *PLOS ONE*, *15*(4), e0230913. https://doi.org/10.1371/journal.pone.0230913

Peelle, J. E., & Davis, M. H. (2012). Neural Oscillations Carry Speech Rhythm through to Comprehension. *Frontiers in Psychology*, *3*(320). https://doi.org/10.3389/fpsyg.2012.00320

Piazza, E. A., Hasenfratz, L., Hasson, U., & Lew-Williams, C. (2020). Infant and Adult Brains Are Coupled to the Dynamics of Natural Communication. *PSYCHOLOGICAL SCIENCE, Vol. 31*(1), 6–17.

Pilley, J. W., & Reid, A. K. (2011). Border collie comprehends object names as verbal referents. *Behavioural Processes*, *86*(2), 184–195. https://doi.org/10.1016/j.beproc.2010.11.007

Plunkett, K., Hu, J.-F., & Cohen, L. B. (2008). Labels can override perceptual categories in early infancy. *Cognition*, *106*(2), 665–681. https://doi.org/10.1016/j.cognition.2007.04.003

Pomiechowska, B., & Gliga, T. (2019). Lexical Acquisition Through Category Matching: 12-Month-Old Infants Associate Words to Visual Categories. *Psychological Science*, *30*(2), 288–299. https://doi.org/10.1177/0956797618817506

Pomper, R., & Saffran, J. R. (2019). Familiar Object Salience Affects Novel Word Learning. *Child Development*, *90*(2), e246–e262. https://doi.org/10.1111/cdev.13053

Pons, F., Lewkowicz, D. J., Soto-Faraco, S., & Sebastián-Gallés, N. (2009). Narrowing of intersensory speech perception in infancy. *Proceedings of the National Academy of Sciences*, *106*(26), 10598–10602. https://doi.org/10.1073/pnas.0904134106

Posner, M. I., Nissen, M. J., & Klein, R. M. (1976). Visual dominance: An information-processing account of its origins and significance. *Psychological Review*, *83*(2), 157–171. https://doi.org/10.1037/0033-295X.83.2.157

Pruden, S. M., Hirsh-Pasek, K., Golinkoff, R. M., & Hennon, E. A. (2006). The Birth of Words: Ten-Month-Olds Learn Words Through Perceptual Salience. *Child Development*, *77*(2), 266–280. https://doi.org/10.1111/j.1467-8624.2006.00869.x

Puccini, D., & Liszkowski, U. (2012). 15-Month-Old Infants Fast Map Words but Not Representational Gestures of Multimodal Labels. *Frontiers in Psychology, 3*(101). https://doi.org/10.3389/fpsyg.2012.00101

Quinn, G. P., & Keough, M. J. (2002). *Experimental design and data analysis for biologists*. Cambridge University Press.

Quinn, P. C. (1987). The categorical representation of visual pattern information by young infants. *Cognition*, *27*(2), 145–179. https://doi.org/10.1016/0010-0277(87)90017-5

Quinn, P. C., & Eimas, P. D. (1996). Perceptual Cues That Permit Categorical Differentiation of Animal Species by Infants. *Journal of Experimental Child Psychology*, *63*(1), 189–211. https://doi.org/10.1006/jecp.1996.0047

R Core Team. (2020). R: A language and environment for statistical computing. *R Found. Stat. Comput. Vienna, Austria*. http://www. R-project. org/., page R Foundation for Statistical Computing.

R Core Team. (2022). R: A language and environment for statistical computing. *R Found. Stat. Comput. Vienna, Austria*. http://www. R-project. org/., page R Foundation for Statistical Computing.

Raviv, L., Lupyan, G., & Green, S. C. (2022). How variability shapes learning and generalization. *Trends in Cognitive Sciences*, *26*(6), 462–483. https://doi.org/10.1016/j.tics.2022.03.007

Riggs, K. J., Mather, E., Hyde, G., & Simpson, A. (2016). Parallels Between Action-Object Mapping and Word-Object Mapping in Young Children. *Cognitive Science*, *40*(4), 992–1006. https://doi.org/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. https://doi.org/10.1111/j.1467-8624.2004.00747.x

Robinson, C. W., & Sloutsky, V. M. (2007). Linguistic Labels and Categorization in Infancy: Do Labels Facilitate or Hinder? *Infancy*, *11*(3), 233–253. https://doi.org/10.1111/j.1532-7078.2007.tb00225.x

Robinson, C. W., & Sloutsky, V. M. (2008). Effects of auditory input in individuation tasks. *Developmental Science*, *11*(6), 869–881. https://doi.org/10.1111/j.1467-7687.2008.00751.x

Rouault, M., & Koechlin, E. (2018). Prefrontal function and cognitive control: From action to language. *Current Opinion in Behavioral Sciences*, *21*, 106–111. https://doi.org/10.1016/j.cobeha.2018.03.008

Saffran, J. R. (2003). Statistical Language Learning: Mechanisms and Constraints. *Current Directions in Psychological Science*, *12*(4), 110–114. https://doi.org/10.1111/1467-8721.01243

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word Segmentation: The Role of Distributional Cues. *Journal of Memory and Language*, *35*(4), 606–621. https://doi.org/10.1006/jmla.1996.0032

Savage-Rumbaugh, E. S., Murphy, J., Sevcik, R., Brakke, K. E., Williams, S. L., Rumbaugh, D. M., & Bates, E. (1993). Language Comprehension in Ape and Child. *Monographs of the Society for Research in Child Development*, *58*(3/4), 1–252. https://www.jstor.org/stable/1166068

Schielzeth, H., & Forstmeier, W. (2009). Conclusions beyond support: Overconfident estimates in mixed models. *Behavioral Ecology*, *20*(2), 416–420. https://doi.org/10.1093/beheco/arn145

Schönbrodt, F. D., Wagenmakers, E.-J., Zehetleitner, M., & Perugini, M. (2017). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods*, *22*(2), 322–339. https://doi.org/10.1037/met0000061

Schöner, G., & Thelen, E. (2006). Using dynamic field theory to rethink infant habituation. *Psychological Review*, *113*(2), 273–299. https://doi.org/10.1037/0033-295X.113.2.273

Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences*, *32*(1), 9–18. https://doi.org/10.1016/j.tins.2008.09.012

Simcock, G., & DeLoache, J. S. (2008). The Effect of Repetition on Infants' Imitation From Picture Books Varying in Iconicity. *Infancy*, *13*(6), 687–697. https://doi.org/10.1080/15250000802459102

Sirois, S., & Jackson, I. R. (2012). Pupil Dilation and Object Permanence in Infants: PUPIL DILATION AND OBJECT PERMANENCE. *Infancy*, *17*(1), 61–78. https://doi.org/10.1111/j.1532-7078.2011.00096.x

Sirois, S., & Mareschal, D. (2002). Models of habituation in infancy. *Trends in Cognitive Sciences*, *6*(7), 293–298.

Slaughter, V., & Suddendorf, T. (2007). Participant loss due to "fussiness" in infant visual paradigms: A review of the last 20 years. *Infant Behavior and Development*, *30*(3), 505–514. https://doi.org/10.1016/j.infbeh.2006.12.006

Sloutsky, V. M., & Fisher, A. V. (2012). Linguistic labels: Conceptual markers or object features? *Journal of Experimental Child Psychology*, *111*(1), 65–86. https://doi.org/10.1016/j.jecp.2011.07.007

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. https://doi.org/10.1111/1467-8624.00570

Sloutsky, V., & Robinson, C. (2008). The Role of Words and Sounds in Infants' Visual Processing: From Overshadowing to Attentional Tuning. *Cognitive Science: A Multidisciplinary Journal*, *32*(2), 342–365. https://doi.org/10.1080/03640210701863495

Smith, L. B. (2005). Emerging Ideas About Categories. In *In L. Gershkoff-Stowe & D. H. Rakison (Eds.), Building object categories in developmental time* (pp. 159–173). LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS. Mahwah, New Jersey.

Smith, L. B., & Thelen, E. (2003). Development as a dynamic system. *Trends in Cognitive Sciences*, *7*(8), 343–348. https://doi.org/10.1016/S1364-6613(03)00156-6

Smithson, M., & Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods*, *11*(1), 54–71. https://doi.org/10.1037/1082-989X.11.1.54

Starr, A. (1977). *Development of Auditory Function in Newborn Infants Revealed by Auditory Brainstem Potentials*. 12.

Stefan, A. M., Gronau, Q. F., Schönbrodt, F. D., & Wagenmakers, E.-J. (2019). A tutorial on Bayes Factor Design Analysis using an informed prior. *Behavior Research Methods*, *51*(3), 1042–1058. https://doi.org/10.3758/s13428-018-01189-8

Suarez-Rivera, C., Smith, L. B., & Yu, C. (2019). Multimodal parent behaviors within joint attention support sustained attention in infants. *Developmental Psychology*, *55*(1), 96–109. https://doi.org/10.1037/dev0000628

Sučević, J., Althaus, N., & Plunkett, K. (2021). The role of labels and motions in infant category learning. *Journal of Experimental Child Psychology*, *205*, 105062. https://doi.org/10.1016/j.jecp.2020.105062

Sweller, J., & Sweller, S. (2006). Natural Information Processing Systems. *Evolutionary Psychology*, *4*(1), 147470490600400. https://doi.org/10.1177/147470490600400135

Swingley, D. (2009). Contributions of infant word learning to language development. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1536), 3617–3632. https://doi.org/10.1098/rstb.2009.0107

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, *76*(2), 147–166. https://doi.org/10.1016/S0010-0277(00)00081-0

Swingley, D., & Aslin, R. N. (2002). Lexical Neighborhoods and the Word-Form Representations of 14-Month-Olds. *Psychological Science*, *13*(5), 480–484. https://doi.org/10.1111/1467-9280.00485

Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition*, *71*(2), 73–108. https://doi.org/10.1016/S0010-0277(99)00021-9

Taxitari, L., Twomey, K. E., Westermann, G., & Mani, N. (2020). The Limits of Infants' Early Word Learning. *Language Learning and Development*, *16*(1), 1–21. https://doi.org/10.1080/15475441.2019.1670184

Terrace, H. S., Petitto, L. A., Sanders, R. J., & Bever, T. G. (1979). Can an Ape Create a Sentence? *SCIENCE*, *206*(4421), 896–902.

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. MIT Press.

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2010). Infant-directed speech facilitates word segmentation. *INFANCY*, *7*(1), 53–71.

Tincoff, R., & Jusczyk, P. W. (1999). Some Beginnings of Word Comprehension in 6-Month-Olds. *Psychological Science*, *10*(2), 172–175. https://doi.org/10.1111/1467-9280.00127

Träuble, B., & Pauen, S. (2007). The role of functional information for infant categorization. *Cognition*, *105*(2), 362–379. https://doi.org/10.1016/j.cognition.2006.10.003

Twomey, K. E., Ma, L., & Westermann, G. (2018). All the Right Noises: Background Variability Helps Early Word Learning. *Cognitive Science*, *42*, 413–438. https://doi.org/10.1111/cogs.12539

Venker, C. E., Bean, A., & Kover, S. T. (2018). Auditory-visual misalignment: A theoretical perspective on vocabulary delays in children with ASD: Auditory-visual misalignment. *Autism Research*, *11*(12), 1621–1628. https://doi.org/10.1002/aur.2038

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

Wang, P. (2004). The limitation of Bayesianism. *Artificial Intelligence*, *158*(1), 97–106. https://doi.org/10.1016/j.artint.2003.09.003

Waxman, S., & Markow, D. (1995). *Words as Invitations to Form Categories: Evidence from 12- to 13-Month-Old Infants*. *Cognitive Psychology*(29), 257–302.

Waxman, S. R., & Braun, I. (2005). Consistent (but not variable) names as invitations to form object categories: New evidence from 12-month-old infants. *Cognition*, *95*(3), B59–B68. https://doi.org/10.1016/j.cognition.2004.09.003

Weisleder, A., & Fernald, A. (2013). Talking to Children Matters: Early Language Experience Strengthens Processing and Builds Vocabulary. *Psychological Science*, *24*(11), 2143–2152. https://doi.org/10.1177/0956797613488145

Weitzman, E. D., & Graziani, L. J. (1968). Maturation and topography of the auditory evoked response of the prematurely born infant. *Developmental Psychobiology*, *1*(2), 79–89. https://doi.org/10.1002/dev.420010203

Werker, J. F., Lloyd, V. L., Cohen, L. B., Casasola, M., & Stager, C. L. (1998). Acquisition of Word-Object Associations by 14-Month-Old Infants. *Developmental Psychology*, *34*(6), 1289–1309.

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*, 49–63.

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

Woodward, A. L., & Markman, E. M. (1998). Early word learning. In *Handbook of child psychology: Volume 2: Cognition, perception, and language.* (pp. 371–420). John Wiley & Sons Inc.

Woodward, A. L., Markman, E. M., & Fitzsimmons, C. M. (1994). *Rapid Word Learning in 13- and 18-Month-Olds*.

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. https://doi.org/10.1080/13506280143000601

Younger, B. (1985). The Segregation of Items into Categories by Ten-Month-Old Infants. *Child Development*, *56*(6), 1574–1583.

Younger, B. A., & Cohen, L. (1986). Developmental Change in Infants' Perception of Correlations among Attributes. *Society for Research in Child Development*, *57*(3), 803–815.

Yu, C., Smith, L. B., & Pereira, A. F. (2008). Grounding Word Learning in Multimodal Sensorimotor Interaction. *In: Proceedings of the 30th Annual Conference of the Cognitive Science Society*, 1017–1022.

Yu, C., Zhang, H., & Smith, L. B. (2006). Learning through multimodal interaction. *In Proceedings of the Fifth International Conference on Development and Learning (ICDL'06).*

Zaadnoordijk, L., Meyer, M., Zaharieva, M., Kemalasari, F., van Pelt, S., & Hunnius, S. (2020). From Movement to Action: An EEG Study into the Emerging Sense of Agency in Early Infancy. *Developmental Cognitive Neuroscience*, *42*. https://doi.org/10.1016/j.dcn.2020.100760

Zhang, F., & Emberson, L. L. (2020). Using pupillometry to investigate predictive processes in infancy. *Infancy*, *25*(6). https://doi.org/10.1111/infa.12358

Declaration of Authorship

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:

Bothe, R., Eiteljoerge, S., Trouillet, L., Elsner, B. & Mani, N. (under review). Better in sync: temporal dynamics explain multisensory word-action-object learning in early development. Registered Report Stage I accepted at *Infancy*.

Bothe, R., Trouillet, L., Elsner, B. & Mani, N. (subm.). Can words do more than arbitrary actions in early category formation tasks? Submitted to *Developmental Science.*

Ricarda Bothe

Göttingen, April 2023

Curriculum Vitae

Ricarda Bothe

EDUCATION

2019 - CURRENT

Georg-August University Goettingen

DOCTORAL STUDENT IN DEVELOPMENTAL PSYCHOLOGY

PhD thesis (working title): "Are words special in early development?", supervised by Prof. Nivedita Mani (Georg-August University Goettingen), Prof. Birgit Elsner (University of Potsdam) & Prof. Alexander Gail (German Primate Center Goettingen), funding awarded by the German Research Association (DFG FG 2253, "Crossing the Borders: the interplay of Language, Cognition and the Brain in Early Human Development").

2017 - 2019

Technical University Kaiserslautern

MSC COGNITIVE SCIENCE

Master's thesis, Title: "The impact of L2 reading direction on the L2 perceptualspan size", supervised by Prof. Shanley Allen and Dr. Leigh Fernandez.

2013 - 2016

University Paul Valéry Montpellier & University of Washington

BSC PSYCHOLOGY

Semester 1- 4: University Paul Valéry Montpellier, France
Semester 5 & 6: University of Washington, Seattle, USA
Bachelor's project, Title: "Complications in pregnancy and autism spectrum disorder development", conducted in cooperation with Seattle Children's Research Institute and supervised by Prof. Sara J. Webb.

EXPERIENCE

03|2023 - current

Georg-August University Goettingen

POSTDOCTORAL RESEARCHER: PSYCHOLOGY OF LANGUAGE GROUP

Using a "Dyadic interaction platform", I investigate the role of curiosity-driven learning in dynamic social contexts

04|2017 - 06|2019

Technical University Kaiserslautern

GRADUATE RESEARCH ASSISTANT: PSYCHOLINGUISTICS & LANGUAGE DEVELOPMENT GROUP, COGNITIVE NEUROSCIENCE & PSYCHOLOGY GROUP

Research topics: "Emerging Grammars in Language Contact Situations: A Comparative Approach", "Language Aptitude: how working memory, grammar abilities, personality, grit characteristics and phonology may facilitate language acquisition", "Investigating flanker compatibility (N2 and ERN)", involved in
EEG data collection & conventional analysis (BrainVision Analyzer),
Eye-tracking data collection (EyeLink 1000) & data analysis (R, lme4),
Audio file transcriptions (PRAAT)

08|2018 - 10|2018

University of Hawai'i at Manoa

GRADUATE RESEARCH ASSISTANT: SECOND LANGUAGE STUDIES DEPARTMENT

Research topic: "Wh-priming in L2 German"
Eyetracking data collection (SMI systems)

08|2015 - 07|2016

Seattle Children's Research Institute

UNDERGRADUATE RESEARCH ASSISTANT: CENTER FOR CHILD, HEALTH, AND DEVELOPMENT

Research topic: Risk factors for autism spectrum disorder
EEG data collection/preprocessing, fMRI data collection
Data management across conducting sites (Yale, UCLA, Boston University)
Evaluation of diagnostic materials

PUBLICATIONS

under review	**Bothe, E.R.,** Trouillet, L., Elsner, B., & Mani, N. (under review). What is special about words in infant category formation? Action and language cues serve similar mechanisms in early object categorization.
under review	**Bothe, E.R.**, Eiteljörge, S.F.V., Trouillet, L., Elsner, B., & Mani, N. (under review). Temporal dynamics in multimodal word-action-object learning in infancy. [accepted registered report, Infancy Journal, January 2022]
2021	Fernandez, L.B., **Bothe, E.R.** & Allen. S.E.M. (2021). The role of L1 reading direction on L2 perceptual span: An eye tracking study investigating Hindi and Urdu. *Second Language Research*, 23. https://doi.org/10.1177/02676583211049742
in prep	**Bothe, E.R.** & Mani, N. (in prep). What's first? On the bidirectional relationship between word- and category learning.
in prep	**Bothe, E.R.** & Mani, N. (in prep). Not the same category? Words promote object category formation in the laboratory, but not online at home.

CONFERENCE TALKS

2022	**Bothe, E.R.** & Mani, N., *How do social interactions guide infant attention and learning?,* Social Curiosity Workshop, Georg-August University Goettingen, Germany, October 2022.
2019	**Bothe, E.R.,** Fernandez, L., & Allen, S., *The impact of L1 reading direction on L2 perceptual span size: investigating speakers of Hindi and Urdu* [Bilingualism], Psycholinguistics in Flanders Conference, Antwerp, Belgium, May 2019.

POSTERS

2022

Bothe, E.R. & Mani, N., *Web-based and laboratory-based infant looking time data?,* Poster presented at Lancaster International Conference on Infant and Early Child Language Development, Lancaster, UK, August 2022.

Bothe, E.R. & Mani, N., *Revisiting the particularity of words in early cognitive development: how words and actions modulate infant categorization behaviour,* Poster presented at International Congress of Infant Studies, Ottawa, Canada, July 2022.

Bothe, E.R. & Mani, N., *Not the same category? Online and laboratory-based infant looking time data,* Poster presented at Workshop for Infant Language Development, San Sebastian, Spain, June 2022.

Bothe, E.R. & Mani, N., *Do words and actions affect object categorization differently across early development?,* Poster presented at Budapest CEU Conference on Cognitive Development, Budapest, Hungary, January 2022. [online]

2021

Bothe, E.R. & Mani, N., *Is it really language? The role on input salience in early categorization,* Poster presented at Lancaster International Conference on Infant and Early Child Language Development, Lancaster, UK, August 2021.

Bothe, E.R. & Mani, N., *Early word learning in cross-domain settings,* Poster presented at Psycholinguistics in Flanders Conference, Kaiserslautern, Germany, May 2021. [online]

Bothe, E.R. & Mani, N., *What is special about words in infant conceptualization? Action and language cues may serve similar mechanisms in early object categorization,* Poster presented at Dubrovnik Conference on Cognitive Science Dubrovnik, Croatia, May 2021. [online]

Bothe, E.R. & Mani, N., *Sorting communication in early category learning: an online study, looking time data,* Poster presented at Budapest CEU Conference on Cognitive Development, Budapest, Hungary, January 2021. [online]

2018

Bothe, E.R., Fernandez, L., & Allen, S., *L1 reading directionas feature influencing L2 Perceptual span size: An eye tracking study.* Poster presented at Fourth Chuo UHM UTokyo Student Conference on Linguistics, Psycholinguistics, Hawai'i, USA, October 2018.

Fernandez, L.B., **Bothe, E.R.**, & Allen, S.E.M., *L1 reading direction plays a role on L2 perceptual span size: An eye tracking study investigating Urdu and Hindi speakers.* Poster presented at Cognitive Neuroscience of Second and Artificial Language Learning Conference (CoNSALL). Bangor, UK, August 2018.

2018 **Bothe, E.R.**, Kolesova, K., Laryea, L., & Fernandez, L., *The influence of L1 reading direction on the L2 perceptual span size.* Poster presented at Psycholinguistics in Flanders Conference, Antwerp, Belgium, June 2018.

2016 **Bothe, E.R.**, Green, A., & Webb, S., J., *Pre-natal and Peri-natal conditions as potential risk factors in the development of autism spectrum disorder.* Poster presented at Mary Gates Symposium, University of Washington, Seattle, USA, May 2016.

TEACHING

2020 - 2023

BSc PSYCHOLOGY

Georg-August University Goettingen

Scientific skills for psychology - Undergraduate seminar and lecture

2020 - 2021

BSc BIOLOGY

Georg-August University Goettingen

Methods in Biocognition - Undergraduate seminar & practical

SUPERVISION

BACHELORS STUDENTS

2022 **Berke Gökbulut**, *Bachelor's thesis*, Title: "Einfluss von frühen Wort-Objekt-Assoziationen auf neue Kategorien", BSc Psychology, Georg-August University Goettingen

2021 **Fiona Berger**, *Bachelor's thesis*, Title: "Verbessert linguistischer Input die Kategorisierungsleistung in verschiedenen Entwicklungsstadien des Spracherwerbs?", BSc Psychology, Georg-August University Goettingen

2020 **Corinna Frank**, *Bachelor's thesis*, Title: "Language and Cognition: How first words foster early categorylearning processes, BSc Biology/Biocognition, Georg-August University Goettingen)

MASTERS STUDENTS

2022 **Sara Oguz**, *Master's thesis*, Title: „Word and action learning in infant category development", MSc Psychology, Georg-August University Goettingen

2021 **Tabea Voß**, *Master's thesis*, Title: „Der Einfluss von Wörtern vs. Handlungen auf das Kategorisierungsverhalten von 1- und 2-Jährigen. Eine Analyse von Webcam-Daten", MSc Psychology, Georg-August University Goettingen

ACADEMIC SERVICE

since 10|2021 Member of the Psychology appeals committee "cognitive modelling"

since 02|2021 Representative of central block / mid-level on institute board

since 09|2020 Student representative of Graduate College "Behaviour and Cognition"

10|2021 Organization Graduate student retreat

TECHNICAL SKILLS

PROGRAMMING LANGUAGES

R, SPSS, Python in training

SOFTWARE

ELAN, GIMP, Adobe (Photoshop, Illustrator), Office (Microsoft Office, Google Docs)

EXPERIMENTAL PROCEDURE

Eyetracking (SMI, EyeLink1000, Tobii, Pupil Labs), reaction time experiments, standardized tests, questionnaires

WORKSHOPS & MENTORING

MENTORING

since 11|2020 Mentee at Margret Maltby Mentoring, program funded by University Medicine Goettingen, mentored by Prov. Dr. Steffi Höhl (University of Vienna).

WORKSHOPS

2021 Mundry, R., *linear models and their application in R,* workshop attended at Georg-August University Goettingen, July 2021. [online]

2020 Mundry, R., *linear models and their application in R,* workshop attended at University of Vienna, February 2020.

2019 Vasishth, S. & Kliegl, R., *Frequentist Stream*, workshop attended at Summer School for Statistical Methods for Linguistics and Psychology, September 2019.

LANGUAGES

German: native

English: fluent (C1)

French: fluent (B2)

Göttingen, 27.4.2023