Title
Meaningful movements in early childhood: The cognitive and developmental bases of dance, gesture, and musical actions

Author
Kim, Min Ju

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Meaningful Movements in Early Childhood: The Cognitive and Developmental Bases of Dance, Gesture, and Musical Actions

A dissertation submitted in partial satisfaction of the requirements

for the degree Doctor of Philosophy

in

Experimental Psychology

by

Min Ju Kim

Committee in charge:

Professor Adena Schachner, Chair
Professor Seana Coulson
Professor Sarah Creel
Professor Gail Heyman
Professor Caren M. Walker

2022

The Dissertation of Min Ju Kim is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

DEDICATION

I dedicate this dissertation to my two honorable grandfathers, Dr. Jong Won Kim and Jeon Koo Lee, who have supported me throughout my journey in graduate school no matter where they were, from earth to heaven.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

NP	Full sample in Experiment 1 who did not receive any prompts to promote or constrain their use of gestures. This participant group was later labeled as No-prompted condition for the comparison with the groups from Experiment 2 (CHAPTER 3)
NP-N	A post hoc group in the No-prompted condition from Experiment 1 who did not produce any gestures during the retelling task (CHAPTER 3)
NP-R	A post hoc group in the No-prompted condition from Experiment 1 who produced gestures including contents that are relevant to the analogous solution in retelling task (CHAPTER 3)
NP-I	A post hoc group in the No-prompted condition from Experiment who produced gestures but only included content that are irrelevant to the analogous solution in retelling task (CHAPTER 3)
SH	Sit-on-hands condition in Experiment 2 (CHAPTER 3)
WG	Watch-gestures condition in Experiment 2 (CHAPTER 3)
WG-N	A post hoc group in the Watch-Gestures (WG) condition who did not produce any gestures during the retelling task (CHAPTER 3)

WG-R	A post hoc group in WG who produced gestures including contents that are relevant to the analogous solution in retelling task (CHAPTER 3)
WG-I	A post hoc group in WG who produced gestures but only included content that are irrelevant to the analogous solution in retelling task (CHAPTER 3)

ACKNOWLEDGEMENTS

I want to thank my advisor Adena Schachner who has taught me to be a true scientist who finds genuine happiness in conducting research and also communicating research with the bigger world. Her true love for music and dance has opened up a new horizon in my scientific identity and my adventure in graduate school was full of joyful grooves than any life could have. I would never forget the moments that we doodled on the huge whiteboard wall in her office together with sparkling ideas, and spending hours in designing a study without getting tired. Also, she is my role model for an academic mama who has the ultimate power to thrive in both work and life the more I get to know her in person. Adena's thoughtful mentorship for my growth as a developmental researcher and her unconditional support for my pursuit of a happy life will always be central to my heart as I proceed to my future steps, being the driving power when I bump into challenges.

I would also like to deeply thank my other committee members – Caren Walker, Gail Heyman, Sarah Creel and Seana Coulson. Everyone was fully supportive of myself replacing one full chapter with another study due to the obstacle of in-person recruitment since the onset of pandemic. I would love to thank them one by one. Despite not being a full student of hers, Caren went out of her way to support me in conducting a first-time correlational study that ambitiously bridges two fields filled with adventures for both of us. Also, Caren was fully supportive of me in achieving cross-cultural research with my native language that I found my other side of academic strength in. Her intellectual insights and warm heart behind the controversial sharks vs. bears ice breakers will always be remembered as I later mentor students short-term or long-term.

Gail was always there for me with moral support and inspiring feedback to all the talks that I had presented since my first year in graduate school. Every conversation I had with Gail I could feel that she truly cares about graduate students and is even eager to learn something new from them. I would be perfectly happy if I could follow just one-tenth of her extraordinary knowledge in the literature, her contagious excitement about true research, and her genuine modesty as a researcher and a mentor. Sarah was the best surprise from the year of SMPC 2017. She is an amazingly charismatic and intellectual scientist that find parallels across multiple areas in research, ranging from early language acquisition to music cognition. Her insights helped me communicate my research to a broader audience with confidence and I will always remember her contagious energy when I want to translate my anxiety in research into excitement. Last but not least, Seana was an invaluable gift from my friends in Cognitive Science department that all deeply appreciated her as a mentor and a researcher. Her stimulating insights with actionable items helped my research shine brighter even from the most fragile moments throughout my dissertation years. I cannot thank her enough for her amazing intellectual support in clarifying my directions in a project that I went out of my safe zone that I grew the most to become a well-rounded researcher. I feel extremely lucky to have met such a grateful collection of brilliant minds throughout my journey in graduate school.

I would love to thank all my collaborators, lab mates, research assistants and colleagues that I wouldn't have been able to reach here without them.

Chapter 1, in full, is a reprint of the material as it appears in Kim & Schachner (2022) in Developmental Psychology. The dissertation author was the first author of this paper. This paper is co-authored by my advisor Adena Schachner. I would also like to thank Madison Pesowski, Laura Cirelli and Haley E. Kragness for the helpful discussion. I would love to thank Diana Deutsch for her invaluable insights for making this project happen from my qualification paper. Thanks also to Kaye Han, Joanna Jain, Angela Mak and Allena McComas for their assistance in data coding and checking. Lastly, I would like to thank the anonymous reviewers, whose valuable advice substantially improved the paper.

In Chapter 2, Experiment 1a has been published in Kim & Schachner (2021) in Proceedings of 43rd Annual Meeting of the Cognitive Science Society. The dissertation author is the first author of this paper. This paper is also co-authored by my advisor Adena Schachner. Thank you to Tim Brady for his insightful and easy to digest classes which taught me to evolve into an experimenter with invaluable skill sets in HTML, CSS and Javascript for Experiment 1a. Thank you to Heidy Davila, Kaye Han, Joanna Jain, Angela Mak, and Alyssa Quy for going beyond their way to make the data collection possible despite the continuity of the pandemic. We also thank Madison Pesowski for her input on the study design especially on Experiment 1b. We also thank Chaolan Lin in her extraordinary help with achieving online data collection in a short time. I would love to thank the UCSD Music Journal Club for providing a stimulating environment for reading papers on music cognition and also for genuine support for this project.

In Chapter 3, Experiment 1 had been published in the proceedings of 43rd Annual Meeting of Cognitive Science Society. The dissertation author was the primary investigator and author of this paper. This paper was co-authored with another intellectual mind, Caren Walker. Thank you to all my hard-working and brilliant research assistants and enthusiastic lab managers that helped this process, Aarthi Popat, Antonia Gruert, Anika Jallorina, Allison Morgan, Emily Songvilay, Jake Troung, Jolie Nguyen, Julia Hoffman, Kristy Kim, Meryem El-Ansari, Nicky

Sullivan, Trisha Katz, and Maddie Weerts. We thank Kensy Cooperrider for his input on the study design, and Jamie Amemiya and Erik Brockbank for statistical guidance. This work was supported by a Hellman Foundation Fellowship and a Jacobs Foundation Fellowship to C. Walker.

Above and beyond, I would like to thank everyone who has supported me throughout this journey in my life beyond my academic life. I would like to thank my husband, Kyu Seo Kim for the unconditional love and support that he has provided me throughout my graduate school. I would like to send unlimited love toward my family in South Korea, my mother, my father, and my brother and both sides of my extended families. I would also love to thank all my friends back in South Korea who have helped me find strength in myself and in my cultural values. I would love to thank all my friends from graduate school that supported each other throughout our endless journey, Alex Rett, Sunyoung Park, Madison Pesowski, Tanushree Agrawal, Elizabeth Bullard, Chaolan Lin, Aubrey Lau, Lauren Oey, Aarthi Popat, Jamie Amemiya, Holley Huey, Elizabeth Lapidow, Elisabeth Marchand, Lim Leong, Jonas Lau, Seo Yeon Jang, Mingi Chung, Danbi Ahn, Suji Jung, Stella Koh, Jean Choi and Sanghyun Lee (among many others). I would also love to thank my family at the UCSD Engaged Teaching Hub at Teaching Learning Commons for providing me a stimulating environment to grow as an educational specialist and also demonstrating a learning community of support. I thank all my fellow graduate teaching consultants, all my mentors (Erilynn Heinrichsen, Carolyn Sandoval and Leah Klement), and my amazing NSF Oral Exam Group with special thanks to Celeste Pilegard and Huihui Qi. I would also love to express gratitude to all the staff members at the UCSD Psychology department for fully supportive of my success as an international graduate student. Last but not the least, I would love to thank my advisor from my Master's program at Yonsei Univeristy, Hyun-joo Song, who is my first role model in academia that I will cherish throughout my life. I thank her for providing me with the most impactful change in my life, pursuing a PhD program in the bigger world outside my hometown to meet all of you.

Finally, I thank all the participants for sharing their time and attention in making this research happen, in-person or remotely on zoom overcoming the limitations during the pandemic. I am grateful to the Birch Aquarium, Fleet Science Center and all of the participating preschools.

VITA

2014 Bachelor of Arts in Psychology, Yonsei University

2016 Master of Arts in Psychology, Yonsei University

2022 Doctor of Philosophy in Experimental Psychology, University of California San Diego

PUBLICATIONS

Kim, M. & Schachner, A. (2022). The Origins of Dance: Characterizing the Development of Infants' Earliest Dance Behavior. Developmental Psychology. DOI: 10.1037/dev0001436.

Moser, C., Lee-Rubin, H., Bainbridge, C.M., Atwood, S., Simson, J., Knox, D., Glowacki, L., Galbarczyk, A., Jasienska, G., Ross, C.T., Neff, M.B., Martin, A., Cirelli, L.K., Trehub, S.E., Song, J., Kim, M., Schachner, A., Vardy, T.A., Atkinson, Q.D., Antfolk, J., Madhivanan, P., Siddaiah, A., Placek, C.D., Salali, G.D., Keestra, S., Singh, M., Collins, S.A., Patton, J.Q., Scaff, C., Stieglitz, J., Moya, C., Sagar, R. R., Wood, B. M., Krasnow, M.M., & Mehr, S.A. (2022). Acoustic regularities in infant-directed vocalizations across cultures. Nature Human Behaviour.

Kim, M., Pilegard, C., Baghdadchi, S., Lubada, M., Delson, N., Sandoval, C., Schurgers, C., Phan, A., & Qi, H. (2022). Leveraging multiple regression to analyze the diagnostic value of adding midterm oral exams for predicting final written exam performance in engineering classes. Accepted for publication in 2022 American Society for Engineering Education (ASEE) Annual Conference & Exposition.

Kim. M., & Schachner, A. (2021). From music to animacy: Causal reasoning links animate agents with musical sounds. In Proceedings of the 43rd Annual Conference of the Cognitive Science Society, Virtual Meeting.

Kim. M., & Walker, C. M. (2021). Preschoolers' spontaneous gesture production predicts analogical transfer. In Proceedings of the 43rd Annual Conference of the Cognitive Science Society, Virtual Meeting.

Bennette, E., Metzinger, A., Lee, M., Ni, J., Nishith, S., Kim, M., & Schachner, A. (2021). Do you see what I see? Children's understanding of perception and physical interaction over video chat. Human Behavior and Emerging Technologies, 1-11.

Kim, M., & Schachner, A. (2020). Gesture Production and Theory of Mind: Effective Disambiguation in Communication through Gesture. In Proceedings of the 42nd Annual Conference of the Cognitive Science Society, Virtual Meeting.

Schachner, A., Kim, M. (2018). Entropy, order, and agency: The cognitive basis of the link between agents and order. In Proceedings for the 40th Annual Conference of the Cognitive Science Society, Madison, WI, United States.

Lee, Y. M., Kim, M. & Song, H.-J. (2017). Socio-emotional cues can help 10-month-olds understand the relationship between others' words and goals. Korean Journal of Child Studies, 38(1), 205-215.

Kim, M. Ahn, E.-J., & Song, H.-J. (2015). Adult Korean EFL learners' pronoun interpretation in discourse context. Journal of Language Sciences, 22(3), 73-92.

Kim, M. & Song, H.-J. (2015). Ten-month-olds' ability to use verbal information when understanding others' goal-directed actions. The Korean Journal of Developmental Psychology, 28(3), 145-158.

Jin, K.-S., Kim, M., & Song, H.-J. (2015). The development of Korean preschoolers' ability to understand transitive sentences using case-markers. The Korean Journal of Developmental Psychology, 28(3), 75-90.

Lee, H-.K., Shin, S.-J., Hwang, M., Kim, M., Lee, H., & Song, H.-J. (2015). A case study of adult good language learners' characteristics in the context of Korea. Modern English Education, 16(3), 185-211.

Ahn, E.-J., Song, H.-J., Kim, M., & Leem, J.-H. (2014). Pronoun resolution in French discourse by Korean learners of French. Korean Journal of Cognitive Science, 25(4), 417-433.

Lee, W.-Y., Kim, M., & Song, H.-J. (2013). The development of Korean children's abilities to use structural cues for sentence comprehension. Korean Journal of Psychology: Development, 26(4), 125-139.

ABSTRACT OF THE DISSERTATION

Meaningful Movements in Early Childhood: The Cognitive and Developmental Bases of Dance, Gesture, and Musical actions

by

Min Ju Kim

Doctor of Philosophy in Experimental Psychology

University of California San Diego, 2022

Professor Adena Schachner, Chair

From the first year of life, children represent others' actions as goal-directed, and can reason about the causes and motivations behind others' observable movements. While the majority of developmental literature has focused on early instrumental goals (e.g., interacting with objects), less attention has been devoted to characterizing children's production and reasoning about non-instrumental movements and non-object-directed goals. However, these kinds of actions have the potential to play a large role in young children's lives in the form of everyday movements such as dance and gesture. In this dissertation, I explore the extent to which children readily produce and comprehend meaningful movements with a rich spectrum of abstract goals. Specifically, I focus on dance-like movement to music, movements that produce musical sounds, and communicative movements that accompany speech (gestures). In Chapter 1, I ask whether infants readily demonstrate dance-like movement to music at home and whether early development of these behaviors has been underestimated by in-lab studies. I show that children produce dance-like behavior earlier than previously believed, and that infant dance shows developmental change with both maturation and learning. I argue that children's dance behavior provides a window into early social, cognitive, and motor development. In Chapter 2, I explore children's reasoning about movements that produce musical sounds. I present evidence that from childhood (6 years of age) onward, rational causal inference plays a role in linking music with agents and movements. I also find evidence of developmental change, such that preschool age children may not engage in causal reasoning about music as older children do. In Chapter 3, I shift to explore the relationship between movement and analogical reasoning. Through both correlational and experimental methods observing children's gesture production, I suggest that children's spontaneous gestures provide a window to their analogical reasoning performance and that enabling them to freely use their hands while thinking helps children actively schematize important structural information over superficial features. Overall, this work shows that children readily produce and reason about socially meaningful movements from early in life and that music, dance, and gesture can serve as unique windows into children's developing cognition.

INTRODUCTION

In daily life, we interpret others' actions to guide our own actions. For example, we look at another driver's hands at a stop sign to drive safely, or we hold the door for the person behind us as we enter a coffee shop. From early in infancy, children also observe other's actions in daily life, from grasping a cup to drink some water to moving to holiday jingles out of pure joy. From these observations, infants readily represent other's movements as goal-directed (e.g., Meltzoff, 1995; Woodward, 1998). For example, when 10-month-old infants see an experimenter reach out their hand toward the same object several times, they infer that the experimenter wants the object as their goal (object-directed goal) rather wanting to repeat the reaching behavior toward the same location (e.g., Lee, Kim & Song, 2017; Woodward, 1998). Even when another's goal is not successfully demonstrated, toddlers can represent what the other human agent intended to do with the objects (Melzoff, 1995). For example, in Melzoff (1995),18-month-old infants observed an experimenter holding a rod with two knobs on each end, then grasping one knob with two fingers, and then demonstrating a pulling-out motion by taking out the fingers to the side. Although the knob did not fall apart from the rod, toddlers successfully imitated what the experimenter intended to do (taking out the knob from the rod), rather than just copying the pulling-out gestures. Such inference was specific to reasoning about human agents and not about mechanical actions from an inanimate machine.

Despite the depth and breadth of goal understanding literature from infancy to childhood, the majority of attention has been devoted to object-directed or instrumental goals (e.g., Baron-Cohen, Leslie & Frith, 1985; Wimmer & Perner, 1983; Woodward, 1998) perhaps due to the need to represent the abstract concept of 'goal' as a physical object in an experimental set-up. Still, there are different layers of goals that are represented throughout daily life, including other non-instrumental goals (e.g., Lakin, Chartrand, & Arkin, 2008). Often, non-object-directed goals include more abstract or higher-level goals such as social affiliation (e.g., waving one's hand to the other person to greet them), self-expression of emotions or preferences (e.g., smiling in response to nice weather, dancing excitedly to one's favorite music), or communication (e.g., using one's hands to give directions to a visitor). These abstract goals carry functions that cannot be solely decomposed into goal objects and rather play unique roles in the early development of social cognition, playing a role in imitation (Meltzoff & Moore, 1977), response to music (e.g., Fujii et al., 2014), and nonverbal communication (e.g., Goldin-Meadow & Wagner, 2005).

According to Novack, Wakefield and Goldin-Meadow (2016), there are three categories of goals that adults can infer from others' actions, or intentional movements: (a) an object-directed goal (e.g., reaching for an object), (b) a movement itself as a goal (e.g., performing a choreographed pattern of dance movements; see also Schachner & Carey, 2013), or (c) a movement that is a representation of a concept (e.g. an iconic gesture). The current work explores reasoning about and production of the last two categories of action, which involve non-object directed goals: movements that are produced for their own sake (particularly rhythmic movement in response to music), and representational movement (particularly, co-speech gestures that represent abstract structural information when solving analogical problems). In addition, I investigate movements that involve objects, but have a non-object-directed, abstract goal: Movements that create musical melodies. While creating music often involves objects like musical instruments (e.g., xylophone keys, guitar strings), these objects serve as the means to an end: The goal is to create an abstract auditory pattern. In addition, the instrumental purpose of music-making is not as clear: Just as movements can be performed for their own sake, music is often produced for its own sake, that is for its aesthetic beauty, intrinsic value, and pure enjoyment. In these ways, actions that produce music (using instruments) are similar to movements that are performed for their own sake (in the absence of objects), and form a category that falls between object-directed goals and movement-based goals.

The final chapter of my dissertation concerns representational movements in early life. People use various media to represent what is on their mind (e.g. *spoken language*, Frank & Goodman, 2012; *sign language*, Erting, Prezioso & Hynes, 1990; *pedagogical actions*, Brand & Shallcross, 2008; *gestures,* Cooperrider, Gentner & Goldin-Meadow, 2016; *dance*, Cirelli, Trehub, & Trainor, 2018). Language is one of the most salient tools for representing thoughts clearly and efficiently (e.g. Frank & Goodman, 2012; Wilkes-Gibbs, & Clark, 1992). Nevertheless, early pre-linguistic infants do not have the full capacity to communicate via language, and therefore display other signals, such as eye-gaze or crying (e.g., Green et al., 1995). Even in early childhood when children can produce longer sentences, often they struggle with choosing the correct words, unintentionally say the opposite word (e.g., saying yes instead of no, Fritzley & Lee, 2003), or have a hard time translating an abstract thought into a full sentence. Therefore, observing representational movements like gesture may capture aspects of children's reasoning that are not expressed in children's speech (e.g., *gestures and mathematical thinking*, Broaders, Cook, Mitchell, & Goldin-Meadow, 2007; *gestures and pragmatics*, Kelly et al, 1999; *dance and social reasoning*, Cirelli, Einarson & Trainor, 2014). In this way, gestures can provide a window into abstract reasoning that broadens developmental perspectives beyond observations from speech (*gestures and abstract thinking*, Cooperrider et al., 2016; *gestures and disambiguation*, Demir-Lira et al., 2018; *cross-modal coordination*, Cornejo et al., 2009).

Therefore, in this dissertation, I present three lines of work that each address *production* of and/or *reasoning* about meaningful movements that are produced in non-object-directed actions in early dance, musical actions, and co-speech gestures to unveil their cognitive and developmental bases. In Chapter 1, I characterize infants' early movement to music, dance. Dance falls under the category of (b) movement-based goal, such that the movement itself is the intended outcome. Dance is a universal human behavior and a crucial component of human musicality (Brown, 1991; Savage et al., 2015). Some studies conducted in labs observed early rhythmic entrainment to music by some infants, but this behavior appeared rare and unchanging over infancy (e.g. Fujii et al., 2014; Zentner & Eerola, 2010). However, anecdotal reports and one case study (Cirelli & Trehub, 2019) suggested that in-lab data could be underestimating the early ontogeny of rhythmic movement in response to music, and that naturalistic at-home data may paint a different picture. Therefore, I characterized infants' earliest dance behavior, leveraging parents' extensive at-home observations of their children. Parents of infants aged 0–24 months (N = 278) were surveyed regarding their child's current and earliest dance behavior (movement by the child, during music, that the parent considered dance), motor development, and their own infant-directed dance. I found that dance begins early: 90% of infants produced recognizable dance by 12.8 months, and the age of onset was not solely a function of motor development. Infants who produced dance did so often, on average almost every day. I also found that dance shows qualitative and quantitative developmental change over the first 2 years, rather than remaining stable. With motor development, age, and more time dancing, infants used a greater variety of movements in dance, and began to incorporate learned, imitated gestures (80% of infants by 17.9 months). 99.8% of parents reported dancing for or with their infants, raising questions about the role of infant-directed dance. These findings provide evidence that the motivation and tendency to move to music appears extremely early and that both learning and maturation lead to qualitative change in dance behavior during the first two years, informing broad questions about the origins of human musicality.

In Chapter 2, I examine how adults and 4- to 6-year-old children *reason about* or *represent* other's movements that produce music, based on their rich concepts of animate agents and inanimate objects. I propose that from childhood, rational causal inference plays a role in linking music with animate agents and movements. Specifically, I hypothesize that when children hear music and see what is causing the music, they will intuitively reason about the physical environment and force that generated the sounds, through a process of Bayesian inference (e.g. Liu et al., 2018). To test this, I asked whether adults and children make flexible inferences about two unobserved variables. First, I ask whether adults and children can infer whether an agent or an object caused musical sounds, integrating information from the sounds' timing and from the visual context in which it was produced (Experiment 1a and 1b). Second, I ask if adults and children can infer what physical environment had to be present based on the sounds they hear, together with whether an agent or object was present to cause the sounds (Study 2). Overall, I find evidence of flexible causal reasoning in adults and 6-year-olds, but observe a potential bias in 4-year-olds to link any orderly musical sounds with animate agents (vs. inanimate objects). I suggest that adults and potentially 6-year-olds are capable of high-level cognition that links music with social concepts, and discuss the developmental trajectory of such flexible reasoning from movements and its outcomes.

Lastly, in Chapter 3, I pivot to focus on representational movement, in the form of co-speech gesture. I test whether gestures *produced* by children can be a window to analogical *reasoning* in a communicative retelling task, where gestures may capture the structural information crucial to solving a new analogous problem. I suggest that observing gesture production in early childhood along with speech can widen the window into early analogical reasoning. Specifically, I asked whether children who spontaneously gesture when completing a retelling task are more likely to engage in analogical transfer, compared to those who do not gesture. To test this, 5-7-year-olds listened to three superficially distinct stories that shared a common abstract problem and solution. After each of the first two exemplar stories, participants were asked to retell the story events to a naïve listener and their speech and spontaneous gesture(s) were coded. For the third story, participants were asked to generate the analogous solution themselves. Results indicate a significant relationship between children's analogical transfer and gesture production. Experiment 1 confirms a positive correlation between children's spontaneous gestures and their analogical transfer success. Experiment 2 added experimental manipulations that either promotes or constrains children's use of gestures. By comparing the three groups of children (no prompts for gesture, encouraged to use hands during the retelling task, and prompted to sit on their hands during the retelling task), I suggest the importance of self-initiated gesture production in capturing the structural information on the child's end and I expand on these observations to the cognitive mechanism behind analogical reasoning.

Overall, this dissertation will shed light on the production and reasoning of meaningful movements in the first few years of life, ranging from commonly-produced movements to abstract stimuli such as music, reasoning about other's movements from contextual cues beyond direct observation, and gestures that are produced in the process of reasoning about structural information. Altogether, I suggest that devoting more attention to how children process and utilize these meaningful movements will unveil a richer spectrum of competence and performance in early cognitive development.

CHAPTER 1. THE ORIGINS OF DANCE: CHARACTERIZING THE DEVELOPMENT OF INFNATS' EARLIEST DANCE BEHAVIOR

Music forms an integral and valued part of our daily lives, impacting well-being (Groarke and Hogan, 2016), social behavior (Kirschner & Tomasello, 2009; Savage et al., 2021), and health (McKinney et al., 1997; Stefano et al. 2004). This is not a unique property of Western, educated, industrialized, rich, or democratic societies (Henrich, Heine & Norenzayan, 2010): Music and dance are universal behaviors, appearing in cultures around the world (Brown, 1991; Savage et al., 2015), and dating back tens of thousands of years into human archeological history (Conard et al., 2009; Nettl, 2015).

How and when does musicality emerge in childhood? It is important to distinguish between music and musicality. "Music" indicates the diverse cultural products generated by and for music-making, such as songs, instruments, and dance styles. In contrast, "musicality" points to the underlying biological capacities that lead us to perceive and produce music – the capacity and natural motivation to engage with music (Honing, 2018). Thus, while music as a cultural product is extremely diverse, culture-dependent and learned, musicality is culturally universal, and its components potentially innate.

Many aspects of music perception are early-developing. From the first weeks of life, song powerfully captures infants' attention and modulates infants' emotions, more effectively even than speech (Corbeil et al., 2016; Ghazban, 2013; Nakata & Trehub, 2004; Trehub, 2011). Infants also respond differently to songs sung in different styles (e.g. lullaby, vs. playsong; Rock et al., 1999), such that songs appear to communicate specific emotional messages even in infancy (Bainbridge et al., 2021; Rock et al., 1999; Trainor et al., 1997; Trehub et al., 1997). Newborn infants can perceive an auditory beat (Winkler et al., 2009), and by 2 months tell apart different tempos and rhythmic patterns (Baruch & Drake, 1997; Demany et al., 1977). These early perceptual sensitivities provide a foundation for learning and musical enculturation, which occurs throughout childhood. For example, through implicit learning of the regularities present in music, children learn the harmonic syntax and rhythmic structures they hear, shaping harmony and rhythm perception in culturally-specific ways (Hannon & Trainor, 2007; Corrigall & Trainor, 2014).

The Developmental Origins of Dance

Dance, particularly rhythmic movement to music, forms a core component of human musicality (Hoeschele, et al., 2015; Honing, 2019; Trevarthen, 1999). In almost every culture, people move rhythmically to music (Brown, 1991; Savage et al., 2015), and in many languages, the concepts of music and dance are referred to using a single word (Trehub et al., 2015). Members of both Western and traditional hunter-gatherer societies relate music to movement in similar ways, suggesting that music and movement share a dynamic structure that supports parallel expression of emotion (Sievers et al., 2013).

What is dance, and how does this behavior develop? Distinguishing between dance as a cultural product, versus the underlying biological capacities that lead people to move rhythmically to music, leads to two distinct questions. When and how does the motivation and tendency to move to music develop? And: How does this behavior change over time, as a process of maturation and learning? If the tendency to move in response to music appears robustly in early infancy, this would suggest that the motivation to move to music is not a product of protracted enculturation. This would have implications for the developmental and cognitive bases of musical engagement, and the link between music and movement (Schachner et al., 2009; Cannon & Patel, 2021). If movement to music is a common part of infants' behavioral repertoire, it would also likely have consequences for early social development: Music and dance positively impact social well-being (Steinberg, Liu & Lense, 2020; Savage et al., 2021), and activities involving movement to music improve parent-child attachment and promote prosocial behavior more than similar non-musical activities (Kirschner & Tomasello, 2010; Vlismas et al., 2013).

How early in life do children produce rhythmic movement to music? Extensive research has focused on entrainment, the ability to move accurately in time with a musical beat (Patel, 2021; Repp, 2005; Schachner et al., 2009). This capacity shows a protracted developmental trajectory: Most children cannot accurately move to a beat until the later preschool years, and do not show adult-like accuracy until age 10 years or later (Drake et al., 2000; Kirschner & Ilari, 2014; Kirschner & Tomasello, 2009; McAuley et al., 2006; Provasi & Bobin-Bègue, 2003). Complex aspects of beat perception, like perception of the hierarchical structure of meters, show a similarly prolonged period of development (Nave-Blodgett, Snyder, & Hannon, 2021). In many ways, however, children show earlier perception of entrained movement than production: By 8 months, infants detect whether others' movements are synchronized with the musical beat (Hannon, Schachner & Nave-Blodgett, 2017), and by one year, infants treat synchrony as socially meaningful, behaving more prosocially toward someone who moved in synchrony with them, compared to someone who moved rhythmically but not in synchrony (Cirelli et al., 2014; Cirelli et al., 2018; Tunçgenç et al., 2015). Young children do show tempo flexibility in movement to music, moving faster for faster tempos than for slower tempos, possibly due to differences in arousal (Cirelli & Trehub, 2019; Kragness et al., 2022a; Yu & Myowa, 2021; Zentner & Eerola, 2010).

Existing lab data show hints of movement to music in infancy, but suggest that this behavior is uncommon and extremely limited. In one influential study, 5-24 month old infants showed more spontaneous rhythmic movement to metrically regular sounds; however, this movement was infrequent, only occurring 6% of the time (Zentner & Eerola, 2010). In addition, the authors report a lack of developmental change in this behavior from 5-24 months (Zentner & Eerola, 2010). In recent work, 6- to 10- month-olds tested in-lab showed equal amounts of movement in silence as to music (de l'Etoile et al., 2020). Below the age of 5 months, only two in a sample of 30 3-4 month old infants produced more movement to music than silence; the vast majority of infants produced less, in effect appearing to freeze in response to music (Fujii et al., 2014). These existing experimental data are consistent with late and protracted production of dance, and suggest that little movement to music occurs during the first two years of life.

However, several factors may lead existing studies to overestimate the age of onset, and underestimate the frequency of dance behavior in infancy. First, infants likely produce more varied and frequent dance in a home setting than in the lab, as children are likely to be more relaxed and comfortable in this familiar setting (Cirelli & Trehub, 2019; Kragness et al., 2022a; Kragness et al., 2022b). Studies of children's early singing have similarly shown better performance at earlier ages from recordings of behavior at home than in the lab (Benetti & Costa-Giomi, 2020; Gudmundsdottir & Trehub, 2018), with recent methods providing evidence of singing in the second year of life (Gudmundsdottir & Trehub, 2018; Reigado & Rodrigues, 2018).

Second, different songs differ dramatically in their tendency to elicit dance behavior in adults, with some songs eliciting dance and others not (Madison, 2006; Madison et al., 2011). Recent work has found that this is true for young children as well: Songs' features and familiarity impact the amount of rhythmic movement children produce (Kragness et al., 2022a; Kragness et al., 2022c). Thus, dispreferred musical stimuli may lead to failure to elicit dance behavior in infants, even for individuals who produce it in other contexts. Many studies of infant movement to music have not customized stimuli to the preferences of different individuals or age groups, instead selecting a small number of standardized tracks as stimuli for all participants (e.g. Zentner & Eerola, 2010; Fujii et al., 2014).

Infants in home settings with self-selected music may thus show a greater frequency and variety of spontaneous dance behaviors, from earlier in life than previously seen in the lab. Anecdotally, this appears to be the case: Parents of young infants appear to commonly report that their infants dance, moving rhythmically in response to certain songs. In addition, a recent case study of a 19-month-old infant provides additional proof-of-concept: At 18-19 months, this child was video recorded at home responding to music with a diverse set of complex, recognizable dance behaviors, with positive emotion; the child's parents reported she began producing recognizable dance behaviors at 6 months of age (Cirelli & Trehub, 2019). Lastly, an exploratory factor analysis of infants' home musical environment found that movement in response to music emerged as a major factor for infants 0-2 years of age (Politimou et al., 2018). Together, these data make plausible the contrasting hypothesis regarding the role of dance in early child development that we test here. Rather than being a rare behavior that is difficult to elicit, and unchanging over the first two years, dance may instead be a common part of infants' daily lives, and may qualitatively change over the first two years of life.

The Current Study

When do infants start producing movement to music, and how does this behavior develop in the first two years of life? We provide a systematic study of the age of onset of children's first recognizable dance behaviors in infancy, and their development during the first two years, with the aim of providing a foundation of empirical data in this novel area as a springboard for future study.

To best characterize infants' first dance behaviors, we aimed to leverage parents' extensive longitudinal observations of their children. We reasoned that observations at a small number of timepoints would be unlikely to capture the very first time an infant produced dance behavior, and thus that short observation sessions (whether at home, or in the lab) would be unable to address our question of interest. Thus, we surveyed parents of infants aged 0-24 months, providing a window into infants' dance behavior over the first two years.

Parents were asked whether or not their child produced dance, and then answered a series of questions regarding their child's dance, motor development, etc. We aimed to lead parents to report any behaviors that were recognizable as dance, regardless of whether they were entrained to the beat or rhythmic. When preschool-aged children are prompted to dance, their behaviors are recognizable as dance to observers, but include both rhythmic and non-rhythmic movements, and are not typically entrained to the beat (Kragness et al., 2022b). We therefore defined the phenomenon of interest for parents as movement by the infant, to any sort of music, that "looks like dancing, to them". Previous findings suggest that people intuitively categorize behavior as dance when the movements appear to be in response to music, and do not serve any other utilitarian purpose (Schachner, 2012; Schachner & Carey, 2013). In doing so, people intuitively use criteria similar to the operational definitions of dance used in experimental studies of animal and child behavior (Keehn, Iversen, Schulz & Patel, 2019; Kragness et al., 2022b) as well as in anthropology and ethnomusicology (Hanna, 1987; Lange, 1975; Royce, 2002). Similar subjective coding has been used to identify dance from children's behavior in previous work (Kragness et al., 2022b).

We focused on several major questions of interest:

(1) How early does the tendency to dance emerge, and how variable is this age of onset across individuals? How frequently or rarely do infants dance? Is infants' dance (like adults') spontaneous and self-motivated, or must it be triggered by others' dance behavior?

(2) Do differences in gross motor development explain differences in the age of onset of dance across children? Or does dance dissociate from motor development: Do some infants begin dancing at early stages of motor development; while others only begin at much later stages? If so, this would suggest that motor development alone is not the sole cause of the onset of dance behavior.

(3) To what extent does infants' dance change over developmental time? Does dance behavior not change substantially between 5 and 24 months of age, as suggested by prior data (Zentner and Eerola, 2010)? Or instead, does dance qualitatively change over the first 2 years (as hypothesized by Cirelli & Trehub, 2019)? With age, and increased motor skill, we predicted that infants would produce more varied dance movements. We also predicted that with age, infants

would begin to include not only simple rhythmic movements, but information that was clearly learned: imitated iconic gestures, as part of action songs (e.g. the movements to 'itsy-bitsy spider'; Trehub, 2019).

(4) In addition, we aimed to characterize the extent to which parents dance with their infants. Few studies have explored parents' dance for their infants, perhaps due to focused interest on the auditory components of infant-parent musical interactions (e.g. Shenfield et al., 2003; Trainor et al., 1997; Trehub et al., 1997, Trehub et al., 2016). Is infant-directed dance rare, or do most parents dance for or with their infants? We hypothesized that infant-directed dance may be a major yet overlooked component of infants' experience.

All data, analysis code, survey materials and supplementary methods and results are available on OSF (https://osf.io/8gqsn/).

Method

Participants

N=278 parents of infants 0 to 24 months old participated (Parents: Mean age= 32.9 years, SD= 4.4 years, Range: 18 to 45 years, 83.1% female, 15.8% male, 1.1% prefer not to answer; Infants: Mean age = 12.9 months, SD= 6.7 months; Range: 0.4 to 24.8 months; 50.4% female, 49.3% male, 0.4% other). This sample size was selected based on the following process and stopping rule: After an initial wave of convenience sampling via social media advertisements (see below), we next aimed to ensure that for each month of age, the sample included data from at least 9-10 participants, through more focused online recruitment (with the exception of 0 months of age, which we reasoned would be particularly difficult to recruit). For any month of age which had more than 9 participants from the initial wave of sampling, these participants were kept in the sample, but no additional recruitment was conducted for that age. This process was selected with the aim of collecting a similar sample size to previous work using parental report to study musical behavior in infancy (Politimou et al., 2018).

Participants were first recruited through advertisements on social media platforms (Facebook, Instagram, and Twitter), and a database of local San Diego families interested in participating in developmental research; these participants volunteered for participation (n=206). To ensure at least 9-10 participants per month of age, additional participants were then recruited by offering $5 payment (on www.prolific.co; n=72; see Supplemental Methods and Results S1 for full age distribution). The study was approved by the Institutional Review Board (IRB), and all participants provided informed consent and all completed the full survey. 5 additional participants completed the full survey but were excluded due to discrepancies in reporting their child's age (2 reported that their child first danced at an age older than their current age; 3 provided ages that differed by 6 months or more when asked to report "age in months" versus "birthdate". For discrepancies below 6 months, we used the birth date to compute age for analyses; see OSF Supplemental Methods and Results S2.1).

Participating parents came from 15 countries and identified as a range of ethnicities (82.7% White, 6.8% Asian, 3.2% More than one race; see OSF Supplemental Methods and Results S1); the majority of participants were currently living in the U.S. (84.5%). Participants from the U.S. came from all four regions of the country (NorthEast, 36.7%; West, 28.6%; South, 24.9%; Midwest, 10.0%; of the 221 out of 235 participants in the U.S. for whom latitude/longitude were available; regions based on U.S. census groupings, U.S. Census Bureau,

2020. See OSF Supplemental Methods and Results S1 for details of approximate locations). Household income spanned a wide range (<$10,000; to $150,000 or more). Most participants had a higher household income than the average family in the U.S.: Of the 79.5% of participants who chose to report income, 46.0% reported a household income of ≥$100,000; 36.6% between $50,000 and $100,000; and 16.1% <$50,000.

Stimuli & Procedure

Participants completed an online questionnaire, which can be viewed on OSF (https://osf.io/8gqsn/). Average time to complete the survey was 12.06 minutes (SD= 8.5), with the exception of 12 outliers who took longer than one hour (Mean + 3SD) and appeared to complete the survey over multiple visits. The survey consisted of six blocks of questions (described below), with blocks presented in the below fixed order.

Child's Demographics

To equate the time required to complete the questionnaire for parents of one vs. multiple children, parents of more than one child were asked to consider only their youngest child in the survey (we reasoned that memories of the youngest child would be most recent, and therefore possibly more accurate). Participants reported how many children they have, and their (only or youngest) child's gender, current age in months, exact birthdate (optional), countries of birth/residency, and whether their child was born full-term or prematurely.

Child's Current Dance Behavior

We first defined the phenomenon of interest for parents as follows: "(a) The child is making the movements: you're not moving for them (like bouncing them to music). (b) They don't have to be "with the beat" or "good" at dancing: just doing something that looks like they’re dancing, to you. (c) You might think a movement is dance if: Your child starts doing it when the music comes on, and then stops doing it when the music turns off. (d) We’re interested in dancing to any sort of music, so including your own singing, or music playing from the radio, etc.” Participants were then asked “Does your child dance? (Yes/No)”. Participants who answered "No" then skipped to the questions on parents’ dance behavior (see below).

Participants who answered “Yes” were asked a series of questions about their child’s recent dance behavior (over the past few weeks), as well as their child’s first dance behavior. These two blocks of questions mirrored one another, and were designed to provide similar information regarding the child’s current and first dance behavior. Parents were asked what is typically happening when their child dances (free-response); about the frequency of dance (How often does your child dance? [1-7 scale; Very rarely to Multiple times a day]); what this dancing looks like (free-response); whether children initiated dance episodes (“Who starts the dancing?”, My child starts dancing first/A parent or other person starts dancing first, and then my child starts dancing too/ Both of these things happen”), and whether children’s dance involved imitating gestures for particular songs, such as the hand motions to Itsy Bitsy Spider, Baby Shark, or any other song (1-5 scale, Never to Always).

Parents were also asked three additional questions, for which results are summarized in the OSF Supplemental Methods and Results S9: How their child is typically positioned while dancing (e.g. lying down, standing up, in a swing), when during the day the child danced (e.g., “During the morning routine”, “During playtime at home”), whether their child had a favorite song, or a particular type of music they like to dance to (free-response), and how often the child’s dance movements appeared to align with the musical beat (1-5 scale, Never to Always).

Child's First Dance Behavior

Participants were then asked to think about the first time they noticed their child dancing, and were prompted to use photos and videos as a memory aid by looking for pictures or videos of their child dancing on their phone or other device. Participants were then asked whether they remembered how old their child was "the first time [they] did something that looked like dancing" (I have a clear idea/I have a guess/I don't remember at all). Those who had a clear idea or guess of their child's age of first dance behavior were asked to provide this age on a slider bar, and rate their confidence in that answer on a 5-point scale. Participants were then asked a set of questions that mirrored the questions regarding children's current dance behavior (described above), but in reference to the child's earliest dance behavior instead of current dance behavior.

Parents' Dance With Child

Participants were then asked about their own dance with or for their child. They were separately asked to consider whether they produced each of three different types of dance: dance with their child in their arms or in a baby carrier (in the arms), dance for the child, so that the child can see them (for the child); and dance together with the child (both parent and child dancing; both together). For each type of parent dance, parents indicated whether they produced this type of dance recently, in the past, or never; and how frequently they produced that type of dance. Parents were also asked when during the day they danced with their child (e.g., "During the morning routine", "During playtime at home"); these results are summarized in the Supplemental Methods and Results S9 on OSF.

Child's Motor Development

To measure children's motor development, we used the gross motor scale from the

Minnesota Infant/Child Development Inventory (Ireton et al., 1977; Ireton, 1992; National Research Council, 2008), a measure that has been extensively validated (e.g., Creighton & Sauve, 1988; Glascoe & Dworkin, 1995; Gottfried, 1984; Guerin & Gottfried, 1987; Sturner et al., 1982). The scale notes specific gross motor abilities (e.g., rolls over back to stomach; walks without help) and asks parents to indicate whether their child does or does not produce each one (Yes/No). We administered the measure in the standard manner: Participants began with questions regarding abilities typical of children half their child's age; the questionnaire ended after parents gave three consecutive "No" responses. The highest motor milestone achieved was thus the highest milestone on which a parent answered "Yes".

Parent Demographics

Participants last provided their own gender, age, country of current residence, country of birth, race and ethnicity, spoken languages, and household income (optional), before completing the survey.

Coding of Free-response Answers

Free-response answers to several questions were coded by two independent coders; the primary coder's data (Coder 1) were used in analyses. On trials where the coders had any disagreement, a third coder reviewed data from the two coders to identify disagreements, determined their own coding of the answer. Coder 1 then revisited their own coding, and was allowed to correct any errors they detected.

Free response answers to "what is typically happening when your child dances?" were coded for whether parents spontaneously mentioned the presence of music. Coder 1 and Coder 2 agreed on whether music was mentioned for 98.6% of participants (Kappa = 0.92); agreement between Coder 3 and Coder 1 was 100%.

Free response answers to “what kinds of movements does/did your child make?” regarding both the first dance and the current dance behavior were coded into descriptive categories. 11 categories were created to capture the different types of movement often described (bouncing/bobbing/nodding, rocking/swaying, waving/flailing, wiggling/shaking/twisting, clapping, squatting/bending knees/twerking, stomping/marching/tapping, kicking, jumping/hopping/kicking, spinning/turning around; any other movement that was not captured in the first 10 categories). Multiple categories could apply to a single answer. Of primary interest was the number of different dance movements produced by infants of each age. Thus, each movement type category was coded numerically, as the number of different movements of that category that were described (0 = no movements of that type; 1 = a single movement of that type; 2 = two independent movements of that type; etc.). These numbers were summed for a total number of different dance movements. Coder 1 and Coder 2 agreed on the total number of different movements produced for 93.2% and 81.7% of infants, for answers regarding first dance and current dance respectively. Kappa was computed separately for each category (of the 11 categories); average Kappa was 0.89 for current dance, and 0.88 for first dance. Agreement between Coder 3 and Coder 1 was 100%.

Free response answers to “what kinds of movements does/did your child make?” were also coded for body parts involved (6 categories; head, arms, hands, hips/waist, body/torso/butt, legs/knees/feet). Because certain types of movement necessarily involve certain body parts (e.g. kicking and jumping involves legs; clapping involves hands), mention of these types of movement resulted in coding of the involved body part even if the body part was not explicitly stated (see Supplemental Methods and Results S8 on OSF). Coder 1 and Coder 2 agreed on the total number of different body parts involved for 91.2% (average Kappa = 0.99) and 98.6% (average Kappa = 0.94) of infants, for answers regarding first dance and current dance respectively. Agreement between Coder 3 and Coder 1 was 100%.

All data, analysis code, survey materials and supplementary methods and results are available on OSF (https://osf.io/8gqsn/). This work was not preregistered.

Results

How Early in Infancy Do Children Begin to Dance?

To determine the age of onset of dance behavior for each child, we drew on data from two relevant questions: (1) Does your child dance [now]?; plotting the proportion of children producing dance at each age; and (2) How old was your child the first time they danced? We further examined whether the dance behaviors in question were initiated by the infant, whether music was present, and predictors of the frequency of dance behavior.

Child's Current Dance Behavior

76.6% of parents (213/ 278) answered that yes, their child dances at their current age (23.4% answered that no, their child does not). To determine the age by which the majority of children dance, we conducted a logistic regression with child's current age as a predictor of dance production. Age significantly predicted whether or not a child danced (Wald Z-test, $\chi^2(1) = 61.4$, $p < 0.001$), with 50% of children dancing by 6.2 months of age, 75% dancing by 9.5 months of age, and 90% dancing by 12.8 months of age (Figure 1.1a).

Retrospective Recall: Age of First Dance Behavior

96.24% of parents whose children dance reported being able to remember the first time their child danced (205 out of 213 parents whose children dance), reporting moderate to high confidence in this memory (M=3.9, SD=0.8; 5-point scale from not confident at all - 1 to extremely confident - 5). The average age of (retrospectively remembered) first dance behavior was 9.4 months, with a standard deviation of 3.8 months, and a range from 0.9 to 20 months (Figure 1.1b). There was a shift toward older ages in the retrospective recall data as compared with the current dance data. This is in part due to differences in the sample: 73/278 participants could not provide a retrospective recall measure of age of first dance, either because their infants did not dance (65), or because they could not remember the age (8). The shift may be further explained by anchoring effects in parents' memory (i.e. anchoring by starting with the child's current age, when recalling the age of first dance; Tversky & Kahneman, 1974).

Overall, these data show that many children dance by 6 months of age, and most children dance within the first year of life. In addition, there is a wide range of age of onset: Some infants produce dance in the first months of life, while others do not do so until the middle of the second year.

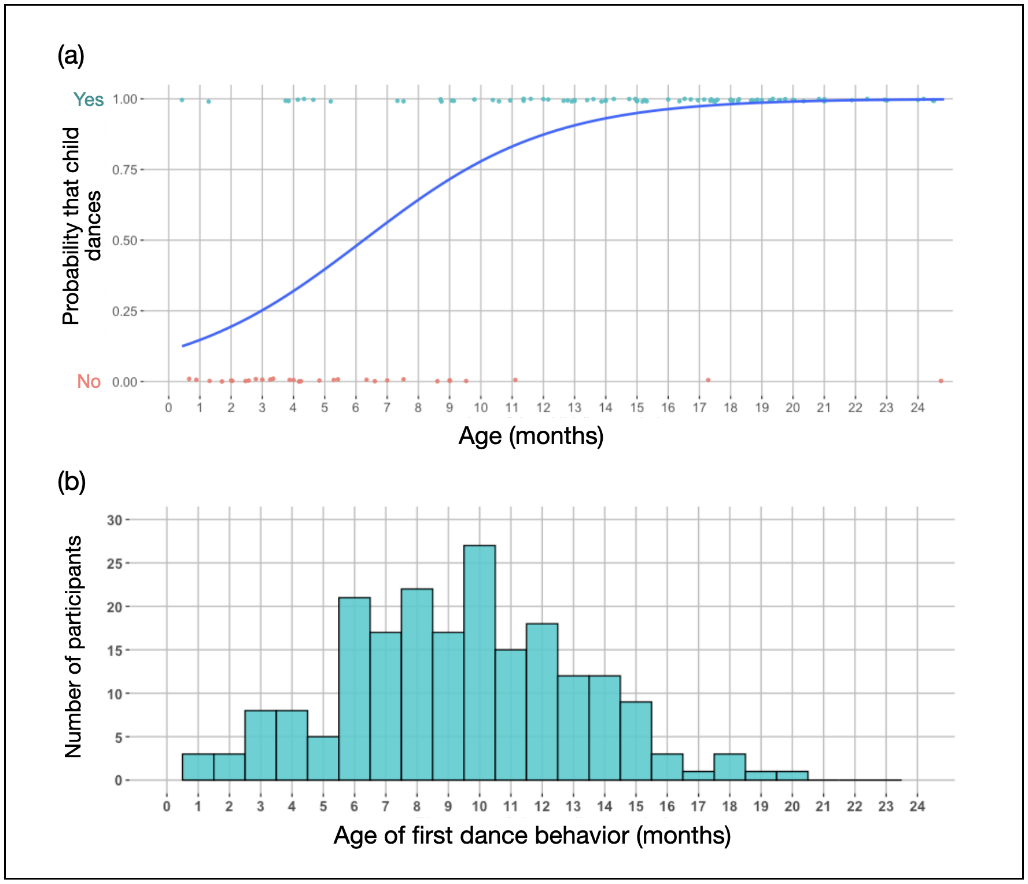

Figure 1.1: Many infants dance by 6 months of age. (a) Current behavior: Parents' reports of whether their infant currently dances, by age (1=yes, 0=no). The blue line shows the fit of a logistic regression. (b) Retrospective recall: Histogram of age of first dance production, based on parents' recall of the age at which this occurred.

Infants Initiate Dance Episodes; Music is Typically Present

Parents reported that their infants' dance behavior was often self-initiated. The vast majority of infants who dance were reported to initiate dance episodes, either sometimes or always (94.4%; 201/213 infants). These infants represented a wide range of ages (age range 0.4-24.8 months, M = 15.3, SD = 5.7; see below for discussion of youngest infants). Only a small number of infants danced only after seeing another adult start dancing (5.6%; 12/213 infants; age range 3.8-22.83 months; Median = 13.0, *M* = 12.6, *SD* = 5.93). Similarly, the vast majority of infants who dance were reported to dance on their own, either sometimes or always (93.9%; 200/213 infants); these infants represented a range of ages (0.4-24.8 months; *M* = 15.3, *SD* = 5.6). Only a small minority of infants danced only in the presence of someone else (6.1%; 13/213; Ages 3.8-22.8 months, *M* = 12.0, *SD* = 5.9).

We also found that music was typically present during the dance behaviors parents identify in our dataset. When asked to describe what is happening when their child is dancing in a free-response format, 91.0% of parents spontaneously mentioned that there was music present (193/213 parents of infants who dance). Only 9.4% of parents (20/213) did not spontaneously mention the presence of music.

Once Infants Start Dancing, They Dance Frequently

Once children first showed dance behavior, they typically danced frequently (mean of 'almost every day', *M* = 5.9, *SD* = 1.4), even if they were under 6 months of age (mean of 'every few days', *M*= 5.1, *SD*= 2.0). With increasing age, children danced more frequently: Age significantly predicted the frequency of dance (linear regression, $F(1,211) = 16.4$, $p<0.0001$); with many infants dancing 'multiple times a day' (88/213 of children who dance, 41.3%).

Infants who Start Dancing Earlier Tend to Dance More Frequently

Do individual differences in the age of onset of dance behavior predict frequency of dance behavior, better than current age alone? To test this, we compared two linear regression models: A simple model including only the child's current age as a predictor of frequency of dance; and a full model including both the child's current age and the age of first dance as predictors. We found that the age of first dance accounted for significantly more variance in children's frequency of dance than current age alone (nested model comparison, $F(1,202) = 5.55$, $p=0.02$).

How Does Motor Development Relate to Age of Onset of Dance?

Motor development significantly predicted children's ability to dance (Wald Z-test, $\chi^2(1) = 34.8$, $p < 0.0001$), with 50% of children dancing when they could 'move forward somehow on their stomach', and 90% dancing by the time they could 'climb up on furniture' (Figure 1.2a). However, this relationship may be due to increasing age, rather than motor development *per se*: Current age also predicts likelihood of dancing (see above), and age was closely related to stage of motor development ($r = 0.91$, $t(276) = 36.7$, $p < 0.001$).

To test whether motor development explained additional variance in children's ability to dance, beyond age, we compared two logistic regression models: A simple model including only the child's current age as a predictor of whether children dance (Yes/No; data from the 'does your child dance now?' question); and a full model including both age and motor development as predictors. We found that compared to the simpler model with just current age as a predictor, the ability to dance was not significantly better predicted when including both motor development and current age as predictors (Nested Wald test, $\chi^2(1) = 2.88$, $p = 0.089$). In contrast, we found that compared to a simpler model with just motor development as a predictor, the ability to dance was significantly better predicted when including both age and motor development as predictors (Nested model comparison, $\chi^2(1) = 5.87$, $p = 0.015$).

This is consistent with the idea that children's age, more so than their motor development, predicts whether and when they begin to produce dance behavior. However, this analysis is made less conclusive because age and motor development were highly correlated. We therefore asked: Does motor development ever dramatically dissociate from the age of onset of dance, at the individual level? Looking at individual children, we find that motor development can dissociate from the ability to dance. Our sample includes four children with advanced gross motor abilities who do not dance (all able to run, climb, and walk without help; motor milestone 18 or above), and six children with very limited gross motor abilities who are reported to dance (unable to roll from back to stomach or sit alone, motor milestone 7 or below; Figure 1.2b).

Are parents of infants with limited motor abilities truly observing music-related, dance-like behavior? These infants are all also very young: below 4 months of age. To determine if parents who claimed their very young infants danced believed the behavior was related to music, we examined parents' free-response answers to the question "what typically is happening when your child dances?" for mentions of music. For five of the eight infants below 4 months of age whose parents observed dance-like behavior, parents stated that the movements occurred in response to music (e.g. "Music plays (any kind) and he moves his head back and forth"). This suggests that some parents of very young infants believe their infants move in response to music. However, we interpret data from very young infants with caution (see Discussion).

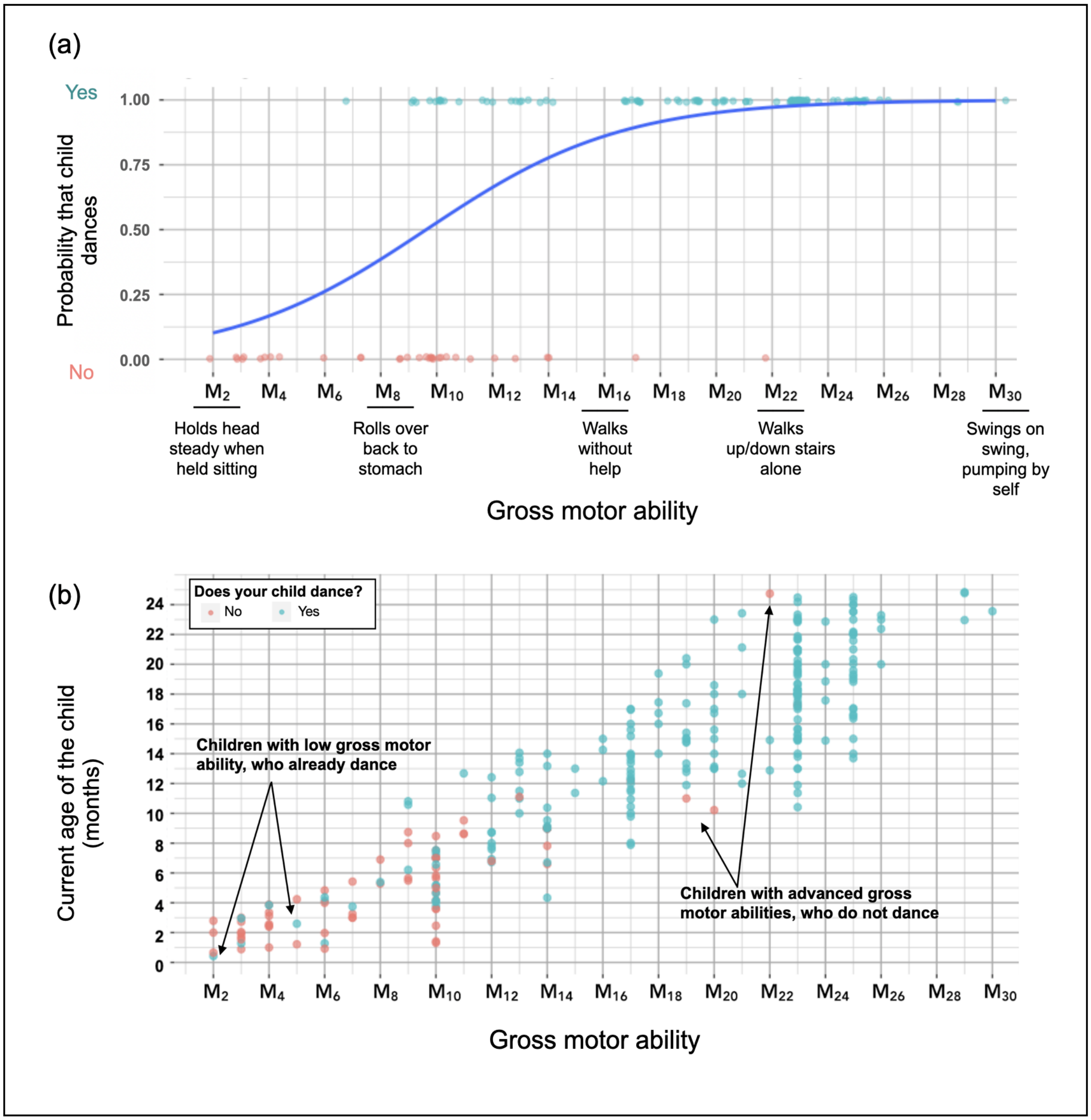

Figure 1.2: Relationship of motor development to production of dance. (a) Gross motor development significantly predicted infants' ability to dance (blue line shows the logistic fit; x-axis shows highest motor milestone achieved; see Supplemental Methods and Results S3 on OSF for all milestone labels). However, motor development did not predict additional variance in ability to dance beyond the predictive power of age alone (see main text). (b) Gross motor development can dissociate from ability to dance: Several infants with limited gross motor ability dance, and several infants with advanced gross motor abilities do not dance.

Does Infants' Dance Change Over the First Two Years?

When asked whether or not their infants' dance movements had changed over time (and if so, how), the majority of parents reported that it had (77.5%, 165/213). This includes parents who explicitly stated 'Yes' (57 out of 165 parents), and parents who responded by describing specific changes to children's dance movements (108 out of 162; e.g. "added specific hand motions"; "starting to involve arms more"; "more controlled and purposeful").

To characterize the rate of change of infants' dance, we analyzed how much developmental time was needed for parents to detect change in a child's dance behavior, for the 205 out of 213 parents who remembered the first age of child's dance behavior. In a logistic regression, amount of time between the age of first dance and the infant's current age significantly predicted whether parents had detected change ($\chi^2(1) = 26.3$, $p < 0.0001$), with 50% of parents detecting change in 1.15 months' duration, 75% in 3.64 months' duration, and 90% in 6.13 months' duration (Figure 1.3).

Older Infants Use Imitated Iconic Gestures in Dance

Age significantly predicted frequency of using imitated iconic gestures in dance, as in action songs such as itsy-bitsy spider ($F(1,211) = 106.0$, $p < 0.0001$, linear regression). This change was not driven by children who do not dance at all, as only parents whose children produce dance answered this question. Children's use of imitated iconic gestures in dance was re-coded as binary (ever, or never); again, age significantly predicted whether dance included imitated iconic gestures (logistic regression, Wald Z-test, $\chi^2(1) = 45.4$, $p < 0.0001$), such that 50% of children produced imitated iconic gestures in dance by 12.5 months of life, and 80% by 17.9 months of life (Figure 1.3b). Imitated iconic gestures were rare in dance before 11 months of age: Of the 46 infants who danced prior to 11 months, only 6 produced imitated iconic gestures, with 3 of these doing so only rarely.

Infants' age, rather than their stage of motor development, predicted whether they produced imitated iconic gestures in dance. A full logistic regression model with both age (as a binary variable) and stage of motor development (maximum motor milestone) was no better than a simple model with only age as a predictor (nested model comparison, $\chi^2(1) = 0.79$, $p = 0.37$); and was significantly better than a simple model with only motor development as a predictor (nested model comparison, $\chi^2(1) = 13.62$, $p < 0.001$).

Greater Variation of Dance Movements is Predicted by Motor Development, Age, and Duration of Dance

To analyze how varied and complex infants' dance movements were, we coded the types of movements and body parts parents mentioned when describing their child's dance (see Methods, Coding of free response answers; and OSF Supplementary Methods and Results S4 and S8). We computed the sum of the coded categories, such that a greater value implies a greater number of movement types or body parts used in dance by that infant.

We first asked if the number of movement types and body parts infants produced changed with age. We analyzed infants' current dance movements (for those infants who produced dance, n=213), and found that with increasing age, infants produced a greater number of different movement types in their dance behavior ($F(1,211) = 32.0$, $p < 0.0001$), though they did not reliably use a greater number of body parts ($F(1,211) = 1.46$, $p = 0.23$).

To understand what factors predict the number of movement types and the number of body parts infants use in their dance behavior, we constructed linear regression models with three predictors: Motor development, age, and the duration of developmental time a child had been dancing (current age - age of first dance). These analyses included all children who produce dance behavior, where age of first dance was reported (n=205). One model predicted the number of different movement types infants produced in dance; the second predicted the number of different body parts infants used.

We found that infants' motor development was a significant predictor of the number of different movement types used in dance (β=0.069, p=0.024), as was the duration of developmental time the child had been dancing (β=0.12; p<0.0001). With these factors included in the model, the child's age was no longer a significant predictor (β =-0.036; p=0.33). Motor development was also a significant predictor of the number of body parts used in dance (β =0.065; p=0.007). Age was also a significant predictor, though this small effect may not be reliable (β =-0.056; p=0.049), and the duration of time dancing was not (β=0.032; p=0.16). Although age and motor development are highly correlated (see above), these analyses suggest that motor development, in addition to age and duration of dance, relates to the use of a greater variety of movement types and/or body parts in dance over the first two years.

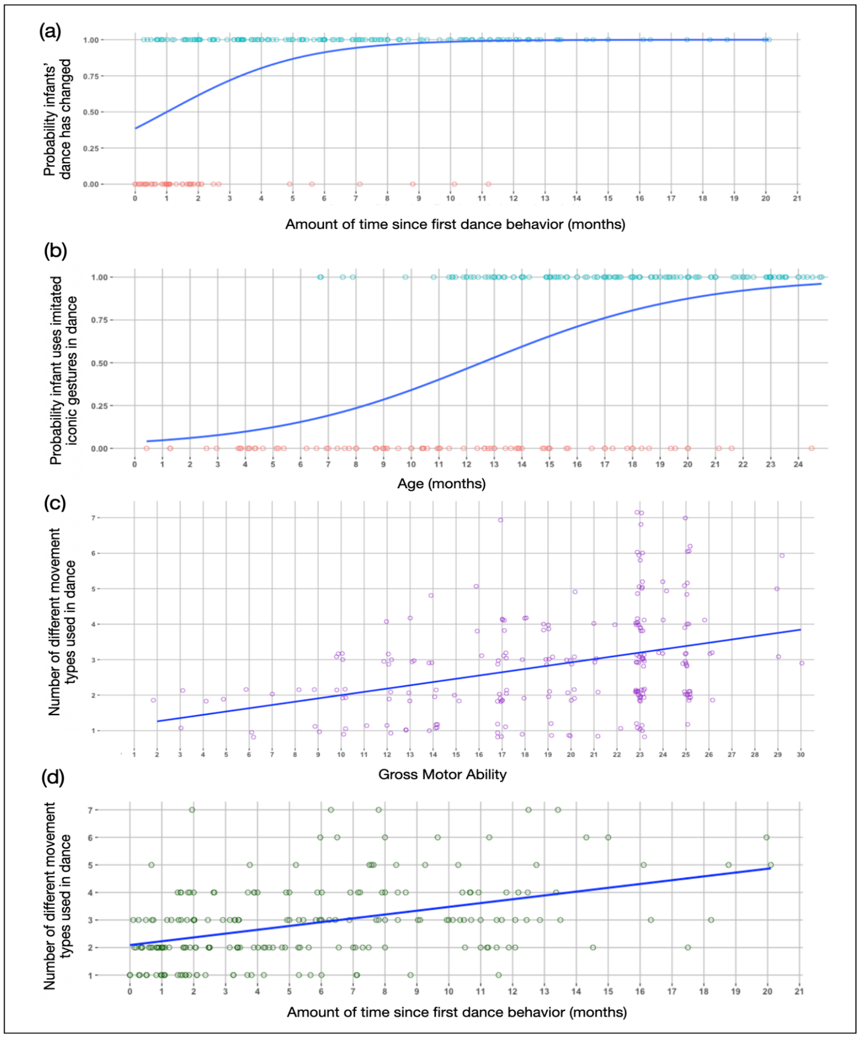

Figure 1.3: The nature of infants' dance changes with time dancing, age, and motor development. (a) The majority of parents reported that the nature of their infants' dance had changed over time, since the time they first danced. This change happened quickly: 50% of parents said dance behavior had changed within 1.15 months; and 90% within 6.13 months of dancing. Blue line shows the fit of a logistic regression. (b) Older but not younger infants produced imitated iconic gestures during dance, e.g. as part of action songs (e.g. itsy-bitsy spider). Blue line shows a logistic regression, with data transformed to a binary measure (Never = 0; All other answers = 1); on the y-axis is the probability that the child produces imitated iconic gestures during dance. (c) With higher gross motor skills, children produced a greater variety of movement types during dance. The blue line shows a linear regression. Movement types were coded from parents' descriptions of their infants' dance behavior (see text). (d) The duration of time that the child had been dancing (current age - first age of dance) predicted the number of different movement types produced by the child in their current dance.

Infant-directed dance: Parents dance for their infants

All but one of the parents in our sample reported that they had danced in the presence of their child over the past few weeks (277 of 278 parents, 99.6%). The only parent who said that they had never danced for their child had a 0.9-month-old infant who did not produce dance. Among the 277 parents who said that they dance with their child, 88.4% had recently danced with their child in their arms or in a baby carrier (245/277); 87.4% had recently danced for their child, with the child watching the parent (242/277); and 81.2% had recently danced together with their child, while their child was also dancing (225/277). Most parents had danced in all three ways, either recently or in the past (76.6%, 213 of 277 parents). Parents reported dancing for their infants relatively frequently, "every few days" on average ('every few days' was 5 on the 7-point-scale; Dancing with child in the arms: M= 5.1, SD= 1.3; dancing for the child, M= 5.1, SD= 1.4; dancing together with the child, M = 5.0, SD= 1.3).

Parents' Dance Changes with the Infants' Age

To test whether parents' infant-directed dance changes depending on the age of the infant, we compared two logistic regression models. A simple model predicted whether or not the parent produced each type of dance recently, with the type of dance (child in the arms, dance for the child, dance together with the child) and the age of the child as predictors. A full model also included the interaction between these two factors as an additional predictor. The full model better predicted parents' dance behavior, such that the kind of dance parents produced changed depending on the age of the child (nested model comparison, $\chi^2(1) = 37.9$, p < 0.001). Specifically, with increasing age fewer parents danced with their child in their arms (Wald Z test for a logistic regression, $\chi^2(1) = 23.7$, $p < 0.0001$). Parents were equally likely to dance for their

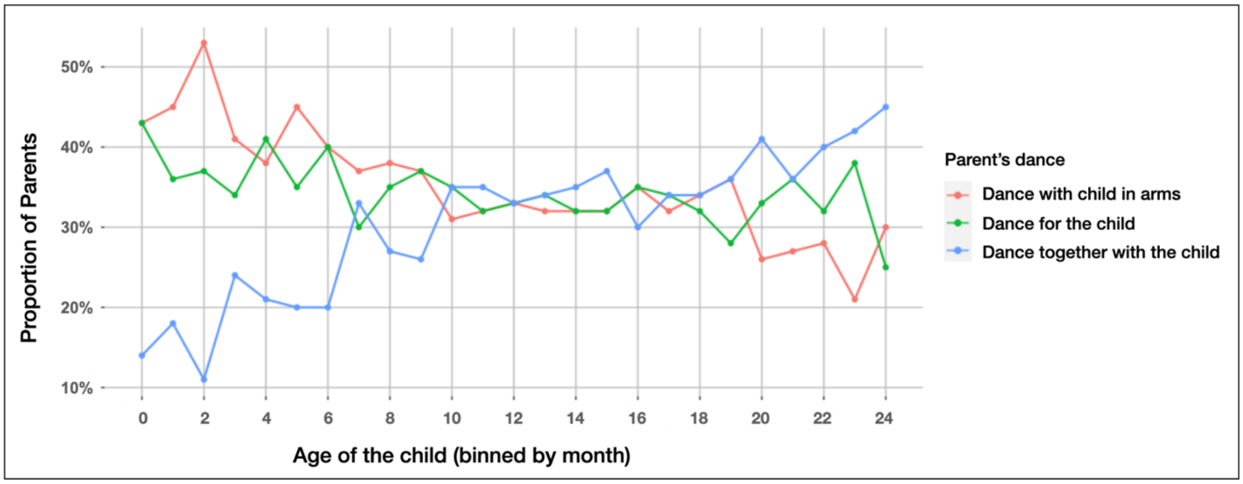

Figure 1.4: Parents' dance changes with the infant's age. Parents indicated whether they dance in the presence of their infant by holding their child in their arms while dancing (red), dancing for the child so that the child can see them (green), and by both the parent and their child dancing together (blue); parents selected multiple answers when applicable. For visualization, the child's age is binned by month; age was a continuous factor in analyses. With increasing age, fewer parents danced with their child in their arms, and more parents danced together with their child while their child also danced. Parents were equally likely to dance for their child at all ages. All points represent the proportion out of all parents who answered that they dance with or to their infants in at least one of the above three ways (n = 277), including parents whose children do not dance.

child at all ages ($\chi^2(1) = 1.9$, $p = 0.16$). With age, more parents danced together with their child while their child also danced ($\chi^2(1) = 44.4$, $p < 0.001$).

Discussion

Overall, we find that in a home environment, movement to music emerges early, is frequently produced, and shows more developmental change over the first two years than previously believed. These findings paint a different picture of early development of dance than do prior lab studies (e.g. de l'Etoile et al, 2020; Fujii et al., 2014; Zentner and Eerola, 2010). At home, with familiar surroundings and music, we find that many infants produce recognizable dance behavior before six months of age (i.e., non-utilitarian, often repetitive, movement in response to music), and a large majority do so within the second half of the first year. Dance is a frequent part of infants' behavioral repertoire, occurring on average almost every day. This suggests that the motivation to move to music is extremely early-developing, and may require relatively little enculturation.

In addition, we find that the nature of infants' dance behavior changes rapidly during the first 24 months of life, such that a majority of parents report noticing change after only 5 weeks. Some of these changes relate to maturation: For example, we find that as motor abilities increase, children produce a greater variety of movement types and use a larger number of body parts in dance, thus producing dance behavior of greater complexity. We also find evidence of learning in the second year of life, and thus evidence of early enculturation processes: By 18 months, over 80% of infants produced clearly learned, imitated gestures as part of their dance actions (as part of 'action songs,' like itsy-bitsy spider). In contrast, infants younger than 11 months rarely did so. This finding suggests that the process of cultural differentiation of dance into its diverse, culturally-specific forms (Honing, 2018) has early developmental roots. Enculturation is known to impact rhythm perception similarly early, in the second half of the first year of life (Hannon & Trainor, 2007). Overall, our data are in line with a developmental story in which the motivation and tendency to move rhythmically in response to music develops early, perhaps spontaneously; and is quickly followed by developmental change due to both maturation and learning, leading to increased variation and complexity of dance behavior over the first two years of life.

Our data also highlight the ubiquity of parents' dance for their infants, which we term infant-directed dance. All but one of the parents in our sample report dancing for or with their infants, doing so on average every few days as a part of daily life. The majority of prior research on infant-parent musical interaction has focused on auditory components of musical behavior, with dance movement either intentionally controlled, or not analyzed (e.g. Shenfield et al., 2003; Trainor et al., 1997; Trehub et al., 1997, Trehub et al., 2016). Current findings suggest that infant-directed dance is a major component of infants' social experience, and supports recent theoretical views positing the importance of multi-modal musical input in infant-parent musical interaction (Trehub, 2019; Trehub & Russo, 2020). In addition, the ubiquity of parents' infant-directed dance raises questions about parallels between infant-directed dance and infant-directed communication in other modalities (e.g. speech, song, gesture; see further discussion below).

The ubiquity of parents' dance also raises the question of whether and how parents' dance relates to the onset of infant dance behavior. We see some evidence of dissociation between parent and infant dance: There are a number of infants who do not dance even though their parents produce infant-directed dance (64 of the 277 infants of parents who produce infant-directed dance). This suggests that observing a parent's dance is not sufficient to trigger the onset of infant dance. However, because all but one of the parents in our sample produce infant-directed dance, our dataset cannot address whether it is necessary: That is, whether any infants dance before they have observed their parents dancing (the single infant whose parents did not dance also did not dance themselves, but was less than one month old). It remains an open question whether it is possible for an infant to begin dancing before they observe dance behavior in others. This an important question for future work, as findings addressing this question would bear on the extent to which infants' early movement to music is spontaneous and self-motivated, and how infants' movement to music relates to relevant aspects of their experience.

Methodological Considerations, and Recommendations for Future Studies

We used parent report as our method of choice in the current work. This method had substantial benefits: By leveraging parents' extensive observations of their children's behavior, we were able to identify the age of onset of infants' dance behavior without the challenge of large-scale longitudinal video recording of years of infants' lives. Collecting large-scale longitudinal video data remains extremely challenging (in terms of logistical and data analytic requirements), and has only been used with small numbers of participants (e.g. 1 participant, recording 24 hours a day; Roy et al., 2015; 3 participants, recording two hours a week; Sullivan et al., 2020). In our study, parent report allowed us to characterize infants' earliest dance behavior across a larger and broader sample than otherwise possible. However, self-selection remains a potential concern: It is possible that people who chose to participate were particularly interested in music and dance, in which case our sample may contain families with higher than average levels of musicality (this is likely both in our main sample, and also in a subsidiary dataset of videos, described below). Future work should probe generalizability, by comparing findings from the current sample to a representative sample of the population, both in the United States, and across cultures (see below for further discussion of potential cultural variation).

Past work has provided evidence for the validity and predictive value of parent report measures of infants' and young children's musical behavior (e.g. Politmolou et al., 2018; Putkinen et al., 2013). Other survey methods provide convergent evidence that parents do not typically overestimate their children's abilities (Oliver et al., 2002; Schneider, Yurovsky, Frank, 2015). However, parent report measures are always potentially limited by observer bias, and by limitations of parents' memory (Seifer et al., 2004). Addressing this limitation is an important avenue for future study. For example, a future study could use novel home recording methods, which have recently proved successful in studying young children's dance behavior (e.g. Kragness et al., 2022a; Kragness et al., 2022b) to obtain video recordings of infants' earliest dance behavior. This would allow for detailed analysis of the movements and their relation to music. Longitudinal follow-up sessions over the first years of life could systematically examine patterns of developmental change, and test multiple key questions raised by the current work.

For example: Do infants' behaviors appear dance-like not only to parents, but also to unbiased outside observers? Recent findings suggest that at least for older infants (1.25 years and above), some behaviors are recognizable as dance to observers (i.e. the video coders, Kragness et al., 2022a). To ask whether any infant behaviors at each of the ages tested here appear dance-like to unbiased observers, we assembled a small subsidiary dataset of video data: 50 video clips of infant dance behavior, two at each month of age from 0-24 months, obtained from systematic searches of a public online video database (www.youtube.com). Full details of the methods and analyses for this subsidiary dataset, as well as full video corpus, are found on OSF. Two coders independently judged whether each of the 50 videos contained behavior that was recognizable as dance (i.e., showed intentional, non-utilitarian, repetitive movement that seemed to be in response to music). All videos of infants aged 4-24 months were judged to include behavior clearly recognizable as dance, by both of the independent coders (100% of both coders' judgements, Kappa = 1, 42 videos). The types of movements seen in these videos were also qualitatively similar to the dance behaviors described by parents in our dataset, and similar to dance behaviors documented in a previous case study of a 19-month-old infant (Cirelli & Trehub, 2019; see Supplemental Methods and Results S11 on OSF for details of similar behavioral features). However, for videos of infants aged 0-3 months, coders disagreed regarding whether the behaviors were recognizable as dance (85% disagreement, Kappa = 0.04, 8 videos). This suggests substantial ambiguity in interpretation of behavior as dance from 0-3 months of life, with this ambiguity decreasing by approximately four months of age.

A small number of parents in our dataset claim that their infants began dancing within the first three months of life. Should these claims be believed? In the first months of life, infants often produce repetitive limb movements, producing such movements far prior to goal-directed motor behaviors such as reaching or crawling (Thelen, 1979, 1981). It is therefore possible that these parents are labeling behaviors as dance that occur during any state of arousal or excitement, rather than being specifically a response to music. However, parents who claim that their very young infants dance cannot be entirely discounted: A previous lab-based study suggested the existence of individual differences at 3-4 months, such that a small minority of very young infants may respond to music with rhythmic movement, even while many others do not (Fujii et al., 2014). Future work should employ prospective methods, or large-scale video recording techniques, to ask what proportion of 0- to 3-month-old infants show dance-like movement in response to music.

By hypothesis, over the first months of life, infants' rhythmic movements may become more specific, shifting from a generic indicator of excitement to a more dance-like form specifically reserved for engagement with music. Infants may move rhythmically more when music is playing than not playing, but also move rhythmically just as much during other high arousal situations – since positive emotion and arousal are typically lower in silence than when listening to preferred music (Zentner & Eerola, 2010; Cirelli & Trehub, 2019; Kragness et al., 2022a). If levels of positive emotion and arousal are matched across conditions, are infants' rates of dance-like behavior higher when music is present vs. absent, and how does this change over development? This future study would provide a test of the above developmental account, by asking whether with age, infants' non-utilitarian rhythmic movements become progressively reserved for times when music is present; and whether their movements during music become increasingly distinct from movements produced when music is not present.

Our findings, in combination with prior literature, lead to two methodological recommendations for future studies. We believe these methodological differences explain why the current dataset finds earlier onset and higher frequency of dance behavior than previous lab-based studies. Firstly, infants appear to produce more varied and frequent dance in a home setting than in the lab, potentially because they are more relaxed and comfortable in this familiar setting (Cirelli & Trehub, 2019; Gudmundsdottir & Trehub, 2018). Thus, studies which aim to elicit spontaneous dance should include controlled observations of behavior at home. Two recent studies provide convergent evidence of extensive dance behavior by young children when recorded at home (Kragness et al., 2022a; Kragness et al., 2022b). Second, because different songs differ dramatically in their tendency to elicit dance behavior in adults, with some songs eliciting dance and others not (Madison, 2006; Madison et al., 2011), dispreferred musical stimuli may lead to failure to elicit dance behavior, even for individuals who produce it in other contexts. Therefore, future studies aimed at eliciting infant dance should make substantial effort to tailor musical stimuli to participants' preferences and age. One recent study of young children has done this, creating customized playlists featuring the favorite song of each individual participant; this may be used as a model for future studies of this type earlier in infancy

(Kragness et al., 2022a). Our open dataset from this study may also include useful information for the selection of stimuli in future studies: Parents in our sample were asked to report specific songs that elicit dance behavior for their infants. These answers may be sorted by age, and used as a database of musical stimuli to elicit dance for infants of particular ages (dataset available in OSF Supplement; summary available in Supplementary Methods and Results S6).

What Explains Variation in Age of Onset of Dance Behavior?

Our data show that children vary widely in the age at which they first dance, with some infants dancing 1.5 years earlier than others. What explains this variation in the age of onset? We find that differences in motor development do not fully explain this variation: Some children did not dance in spite of advanced motor development, while others danced from very early stages of motor development.

One possibility is that differences in age of onset of dance behavior reflect stable, trait-like individual differences in musicality, or motivation to engage with music. In this case, infants who danced earlier would be expected to be more motivated to engage with music than those who dance later, and these levels of motivation would remain stable over time. In line with this prediction, infants in our dataset who began dancing earlier in infancy also tended to dance more frequently later in toddlerhood, as compared to those who started dancing later in infancy. It remains an open question whether variation in dance behavior in infancy predicts individual differences in musical engagement into later childhood or adulthood. Future longitudinal work may test this prediction.

If so, this would support the hypothesis that early variation in tendency to dance in infancy is an early-developing sign of stable, trait-like individual differences in motivation to engage with music -- driven by both genetic and experiential factors[1]. This would be in line with adult data: Adults show individual differences in their tendency to engage in dance, with multifactorial causes (musicality, emotional expressivity) that do not reduce to differences in motor ability (Hoeschele et al., 2015; Müllensiefen et al., 2014; Patel 2008; Iversen & Patel, 2008), and behavioral genetics suggests the existence of heritable factors affecting individual levels of musicality (Bachner-Melman et al., 2005; Gingras et al., 2015).

Infant-directed Dance By Parents

We also found that infant-directed dance by parents is an extremely common phenomenon, such that all but one of the parents in our sample danced for or with their infants. This supports the idea that infant-directed dance may be an important part of infant social experience. In addition, the way that parents danced with their children changed with age, raising the possibility that parents may modify their infant-directed dance to be developmentally appropriate.

Adults around the world speak, sing, and gesture to infants in particular ways, modifying their behavior systematically in ways that maximize infant attention and learning. For example, infant-directed speech and song are more repetitive, simpler, slower, higher in pitch, have greater pitch jitter or pitch contour, and have overall higher emotionality (Cooper & Aslin, 1990;

[1] Although all but one of our parents in our sample danced for their infants, parents likely differ in the amount they dance to and listen to music, with these differences in parent behavior likely consistent and stable over time (e.g. Mehr, 2014; Politimou et al., 2018). This would result in more or less musical exposure, and thus fewer or greater opportunities for dance, across individual infants. These individual differences may also be partially driven by genetic factors: As most of the infants and parents in our sample are likely biologically related, correlated differences between parents' and infants' musical engagement may also be caused by shared genetics.

Fernald, 1992; Trainor et al., 2000; Trehub & Nakata, 2001; Trainor et al., 1997). Infant-directed speech and song are more effective at holding infants' attention than are adult-directed speech and song (Cooper & Aslin, 1990; ManyBabies Consortium, 2020; Trainor, 1996), and infant-directed acoustic features promote infants' learning (Eaves et al., 2016; Kaplan et al., 1996; Thiessen et al., 2005). Infant-directed modifications are not solely acoustic phenomena: Gestures are modified in parallel ways, to be more repetitive, simple, and have higher emotionality when directed at infants (Brand et al., 2002). Like infant-directed speech and song, infant-directed gestures are more effective at holding infants' attention than adult-directed gestures (Brand & Shallcross, 2008).

Infant-directed dance may have similar properties: Parents may modify the way they dance to optimize infant attention and learning, and in doing so incorporate properties that parallel the features of infant-directed communication in other modalities. Future work may explore the properties of infant-directed dance, and test the hypotheses that (a) parents produce dance movements that are modified in ways that parallel the features of other infant-directed communication, and (b) these modifications serve to maximize infants' attention and learning.

Cross-cultural Comparisons in Early Development of Dance

What role do cultural factors play in shaping infants' motivation and tendency to dance? Although our sample contained some diversity in ethnicity and country of origin (see OSF Supplemental Methods and Results S1), the majority of our participants were white and from the United States. We point to cross-cultural comparisons as an important avenue for future work.

Dance behavior may begin in a universally similar way in early infancy, and then differentiate by culture later in infancy and childhood. Alternatively, even the early development of dance may differ cross-culturally, depending on cultural norms and practices, and differences in musical exposure. There are relevant cultural differences early in infancy: For example, cultures differ in the extent to which infants are swaddled, and/or worn by an adult caregiver in a wrap or baby carrier (Little, Legare & Carver, 2019). This variation would be expected to impact infants' exposure to rhythmic movements (e.g. while carried); as well as the ability to demonstrate observable early movement to music (e.g., such movements may be less easily produced or observed when swaddled).

There are also marked cultural differences in the acceptability and value of dance (e.g. Brazil, where dance is valued; vs. Iran, where dancing may be prosecuted; and Amish communities, where dancing and instrumental music are considered immodest and are rare). Relatedly, there are marked cultural differences in the acceptability of openly expressing emotions: Emotional restraint is considered desirable as a form of self-control in collectivistic (e.g. East Asian) cultures, but not in individualistic (e.g. American) cultures (Matsumoto et al., 2008). Emotional restraint is valued differently across cultures in children as well as adults (Louie et al., 2014). These differing norms may impact the acceptability of emotional expression through dance from early ages. Future cross-cultural work should probe cultural differences and similarities in early dance behavior, including asking when and how differing cultural norms and childcare practices impact infant dance behavior.

Conclusion

Overall, here we provide evidence that production of musical behavior, in the form of dance, is common in early infancy and is a frequent part of infants' behavioral repertoire. Over the first two years, dance behavior also shows qualitative developmental change, rather than remaining stable as previously believed. We find that with age, infants dance movements become more varied; and that developmental change in the first two years of life reflects both maturation and learning. Further, we find that parents commonly dance for their infants. Overall, our findings paint a novel picture of early dance as a key part of infants' behavioral repertoire; and provide a foundation for understanding movement to music as a fundamental component of infants' cognitive, emotional, social and motor development.

Acknowledgements

Chapter 1, in full, is a reprint of the material as it appears in Kim & Schachner (2022) in Developmental Psychology. The dissertation author was the first author of this paper.

I would like to thank Sarah Creel, Seana Coulson, Gail Heyman, Caren Walker, Madison Pesowski, Laura Cirelli and Haley E. Kragness for helpful discussion. Thanks also to Kaye Han, Joanna Jain, Angela Mak, Allena McComas and Alyssa Quy for their assistance in data coding and checking. Lastly, we are all greatful to all the parents who have participated in this research. The authors have no known conflict of interest to disclose.

CHAPTER 2. FROM MUSIC TO ANIMACY: CAUSAL REASONING LINKS MUSICAL SOUNDS WITH ANIMATE AGENTS

The mere act of listening to music activates more than just auditory representations. Music also activates representations of movement, even when no movement is produced or viewed (Cannon & Patel, 2021; Chen, Penhune & Zatorre, 2008; Grahn & Rowe, 2009). Music appears to activate social representations as well, particularly for music that people believe was created by a person (Launay, 2015; Steinbeis & Koelsch, 2009). Why would musical sounds make us think about movement and social agents?

At least in part, musical sounds are linked with movement through low-level, modality-specific connections: Features of rhythmic sound drive the motor system (and vice versa) via early-developing connections between auditory and motor systems (Cannon & Patel, 2021; Chen et al., 2008; Grahn & Rowe, 2009; Patel & Iversen, 2014). Beyond rhythm, other auditory features also drive perceptions of animacy, movement, and emotion in music (Blust, Baker, Richard & Shanahan, 2016; Juslin & Laukka, 2003; Sievers, Polansky, Casey & Wheatley, 2013). In particular, the organized nature of musical sound may lead listeners to think about agents since people learn to associate orderly objects and sounds with agents early in life (Ma & Xu, 2013; Newman, Keil, Kuhlmeier, & Wynn, 2010; but see Schachner & Kim, 2018). Numerous studies have focused on the role of perceptual attributes of sound in driving links with movement and social concepts.

Here we propose that an additional cognitive process may also play a role in this linkage. In particular, high-level causal reasoning about how the music was generated may lead people to link music with movement and agents through the process of drawing inferences about the best explanation (Keil & Newman, 2015; Lipton, 2004; Ma & Xu, 2013; Tenenbaum, Griffiths & Kemp, 2006). Causal reasoning has the potential to explain why music is linked both to representations of movement as well as to animate agents. The linkage may result from a joint inference (Baker, Jara-Ettinger, Saxe & Tenenbaum, 2017) about the type of movement required and the type of force that could cause those movements. According to this account, to explain an observed sequence of musical pitches, people infer what types of movements are required to produce them, and they simultaneously infer the most likely cause of those movements, weighing multiple hypotheses (e.g., perhaps they were generated by a person; by an animal; by gravity; by the wind; etc.). They then choose the hypothesis that is both plausible a priori and provides a good explanation of the observed events.

This type of inferential reasoning occurs in multiple related domains, including physical-mechanical reasoning, social reasoning, and mental state inference (Teglas et al., 2011; Baker et al., 2009; Battaglia, Hamrick & Tenenbaum, 2013). Here, we ask whether causal reasoning, which is a domain-general ability, also occurs when listening to musical sounds, and how this capacity develops in childhood. To be clear, we recognize that causal reasoning is not unique to musical sounds and that the causal reasoning process likely occurs in a similar way for non-musical environmental sounds (Gerstenberg, Seigel & Tenenbaum, 2018; Traer, Norman-Haignere & McDermott, 2020). However, if people engage in causal reasoning about musical sounds in particular, this would provide a parsimonious explanation for several aspects of musical representations.

Causal Reasoning, and Animate Agents as Causes of Music

Causal inference should allow adults, and even young children, to infer that many musical sounds are caused by animate agents. From early in life, people have a mental model of the physical-mechanical world that allows them to predict and explain physical-mechanical events; they have a distinct mental model of animate agents as differing from inanimate objects (Battaglia, Hamrick & Tenenbaum, 2013; Spelke & Kinzler, 2007). In particular, animate agents have unique capacities, including self-propelled, goal-directed movement (Gelman, Durgin, & Kaufman, 1995; Gergely et al., 1995). Thus, when we see an event that requires self-propelled movement, we infer that an animate agent is the cause (e.g., Saxe, Tenenbaum & Carey, 2005).

Similarly, when we hear sounds that require self-propelled movement, we may infer that an animate agent is the cause. Existing work shows that adults use rational causal inference to reason about the sources of non-musical sounds (Gerstenberg et al., 2018; Traer et al., 2020). For example, when faced with a pinball-machine-like box into which a ball could be dropped, participants were able to use the timing of impact sounds to infer the path taken by the ball (e.g., where it entered the box). Participants were also able to use auditory information in conjunction with visual context to predict where the ball was likely to land. In both cases, participants' judgements were well-predicted by a Bayesian model of rational causal inference (Gerstenberg et al., 2018). Gerstenberg et al. (2018) have shown that people can use properties of sound to infer physical-mechanical information, such as an objects' path of movement. However, they did not test whether people can make the more abstract inference that certain sounds were produced by animate agents, or how the presence vs. absence of agents impacts inferences about what occurred to produce musical sounds, questions which are explored in the present study. We expect that people will be able to make such inferences by using causal inference in a similar way for musical sounds as for non-musical environmental sounds. Our focus is on musical sounds (specifically, short melodies played on a xylophone) for two reasons. First, causal reasoning about abstract, instrumental musical sounds would have implications for the nature of musical representations, helping to explain why these sounds activate representations of movement (Grahn & Rowe, 2009) and social concepts (Launay, 2015; Steinbeis & Koelsch, 2009). Second, short instrumental sequences can be produced by people or inanimate forces (e.g., wind chimes), leading to uncertainty that may be resolved with causal inference. This situation allows us to detect whether people engage in this reasoning when hearing musical sounds.

One signature of causal inferential reasoning is "explaining away"—the idea that a plausible alternative explanation weakens evidence for other explanations. This phenomenon is a signature of causal inference, and it serves as evidence that the underlying cognitive representation is rich and structured (Gopnik & Sobel, 2000; Pearl, 2000; Tenenbaum et al., 2006). If causal reasoning can link musical sounds with agents, then the presence of a plausible physical-mechanical cause could "explain away" the agent, reducing the likelihood of inferring an agent is the cause, even for the same musical stimulus. Furthermore, such causal reasoning should be also observed with different sets of evidence – when the the agent or the inanimate object is present but the physical-mechanical cause is absent, such inference would reflect that agents can produce the same musical sequence on any physical-mechanical environment but the inanimate object would be constrained to the environment in its outcomes.

In previous work, we explored a similar topic—why orderly stimuli are associated with animate agents—by creating an experimental paradigm that may be useful for investigating the role of causal inference in music perception (Schachner & Kim, 2018). In particular, participants were shown a xylophone-like staircase, in which the bars of the xylophone were embedded in the steps. In one condition, the bars were arranged so that a ball rolling downhill due to gravity would produce a descending scale (G-F-E-D-C). In another condition, the bars were placed in a different position such that the ball would play the same notes in a different order as it rolled downhill. The xylophone was occluded from view, and either a descending scale or the same notes in a different order were played. Participants were asked to judge if either an inanimate object (a ball) or an animate agent (a cartoon character) had been present when the sounds occurred. Participants reliably inferred that an animate agent was present only when the sounds and placement of the bars on the xylophone implied that the character would have had to change direction or jump (Schachner & Kim, 2018).

These data suggest that people may use pitch information (the order in which pitches are produced) to infer whether an agent caused musical sounds. However, in that study, it was possible for participants to solve the task without using auditory information. Because we were not concerned with isolating the role of auditory information, the task also provided a visual "legend" of numbers corresponding to each pitch as notes were played to aid those experiencing difficulties with pitch perception or memory (Schachner & Kim, 2018). Thus, participants could have solved this task without reasoning about the causes of the musical sounds by relying solely on visual cues.

Development of Causal Reasoning

In addition to exploring adult causal reasoning about musical sounds, in the current work we ask how this capacity develops. Do children spontaneously reason about the causes of musical sounds? There are several reasons to expect that infants and young children may be able to infer that agents are the causes of the musical sounds they hear. From infancy, children differentiate between animate agents and inanimate objects and know that agents have the unique capacity for self-propelled, goal-directed movement (Spelke & Kinzler, 2007; Gelman et al., 1995; Gergely et al., 1995). When infants see an event that requires self-propelled movement, they infer that an animate agent must be the cause (Saxe et al., 2005). In the first months of life, infants link sounds with specific events, for example, expecting to hear a sound when they see an impact between two objects or between an object and a surface (Bahrick, 1983, 1988; Kopp, 2014; Lewkowicz, 1992, 1994; Spelke, 1979). Thus, even infants may engage in causal reasoning about sounds and realize that certain sounds could not be produced without an agent's intervention.

Alternatively, this reasoning may undergo developmental change during the pre-school years, and not all children might display adult-like level of reasoning. There are several possibilities regarding why and how younger children may differ from adults. First, previous studies have shown that infants and young children exhibit a bias: They expect that orderly things were created by animate agents (Newman et al., 2010; 2015; Ma & Xu, 2013). Different musical sequences can differ in orderliness: For example, a descending scale is highly orderly, while a scrambled sequence of the same tones is perceived as less orderly (Schachner & Kim, 2018). Similarly, a five-note melody where the notes are evenly timed throughout the sequence is more orderly than the same five-note melody where there are uneven pauses between the notes. If young children do not use causal reasoning, but instead expect that more orderly sound patterns are more likely to be created by agents, then children may judge any musical outcome that sounds orderly (vs. disorderly) to have been created by the agent, regardless of other factors (e.g. the environmental context that the music was created, and whether self-propelled movement would have been required to produce it). Alternatively, children might be biased to link *all* musical stimuli with animate agents. Children may tend to observe an association between musical production and people, and thereby through associative learning link any and all musical sounds with agents (Launay, 2015).

Lastly, younger and older children could differ due to failures to integrate visual with auditory information. Reasoning about the causes of sounds requires pulling together information from these two modalities. However, multimodal integration has been found to develop slowly over childhood, with adultlike integration not occurring until age 8-10 (Barutchu et al., 2009; Gori et al., 2008; Nardini et al., 2008). Young children also fail to make some nuanced sound source judgements that adults easily make, particularly regarding whether water is hot or cold from the sound of it being poured (Agrawal & Schachner, 2022). In a recent developmental study of causal reasoning about non-musical sounds (which also required auditory-visual integration), children before age 6 did not engage in causal reasoning about the movements required to cause the sounds, and did not integrate vision and sound; by age 6-8 years, many children did (Outa, Zhou, Gweon, Gerstenberg, 2022).

These data suggest that children may only show adult-like causal reasoning about how musical sounds were generated at age 6 years, but not before. At younger ages, these data make the further prediction that children may either link orderly musical sounds with agents, or may link all musical sounds with agents.

The current work

Therefore, several experiments were conducted to examine causal reasoning about musical sounds in both adults and children. In Experiment 1, we ask whether adults (Experiment 1a) and children (Experiment 1b) can use information from musical sounds to infer whether the sounds were caused by an agent. We focus on the use of timing in musical sounds, as past work shows that adults and older children can make physical-mechanical inferences based on the timing of non-musical sounds (Gerstenberg et al., 2018; Outa et al., 2022). Can people integrate auditory information about musical sounds with visual information about the environment to rationally infer whether an agent or object was the most likely cause of a musical sequence? If people use their rich mental models of the physical and social world to reason about musical stimuli, they should be able to infer whether musical sounds were caused by an agent by making a joint inference about the movements required to cause the sounds and the type of agent or object required to cause those movements. That is, when a sound sequence requires self-propelled movement, people should infer it was caused by an agent. Alternatively, when a sound sequence could be produced by inanimate forces (e.g., gravity), people should "explain away" the agent, making them less likely to infer that an agent caused the sounds.

If people have a mental model of the causal process, then they should be able to flexibly infer other components of the causal system, such as the structure of the environment. In Experiment 2, we therefore explored a related inference, asking if adults (2a) and children (2b) can infer the physical environment in which the music was created based on the musical notes they hear, and whether an agent or object was present. If people can use causal reasoning to infer the physical environment in which the music was created, then people should be more uncertain about the nature of the environment when the agent is present. This is because the animate agent can produce many different paths of movement, and thereby produce many different musical sequences. In contrast, when only an object is present, people should be more certain, choosing the environment that would allow the observed sounds to be produced by inanimate forces alone (e.g. a ball rolling downhill).

Experiment 1a

When people listen to music, can they flexibly reason about how the musical sounds are generated in a given physical environment? Accounts of inverse planning (e.g., Liu et al., 2018) suggest that people intuitively reason about the causes of observed events, ranging from social (e.g., Newman et al., 2015) to physical-mechanical events (e.g., Gerstenberg, 2018). In doing so, rich concepts of social agents may support or hinder their causal reasoning process: People might flexibly reason that an animate agent could have produced self-propelled movements to create music (e.g., jumping between xylophone bars to create a musical melody) or people may be biased to associate all orderly events (e.g., a descending scale) with the existence of an animate agent (vs. an inanimate object). To test whether people integrate auditory and visual information to infer whether an agent was the cause of musical sounds, we adapted prior methods (Schachner & Kim, 2018) to design animated stimuli in which xylophone-like musical bars were embedded in a hill. In particular, we manipulated the *visual context* by changing the distance between the bars, either placing them at even distances from one another (evenly-spaced-bars xylophone) or in unevenly-spaced groupings (unevenly-spaced-bars xylophone; see Figure 2.1).

In each trial, the visual context was occluded, a sound sequence was played, and participants were asked to judge whether an animate agent or an inanimate object had been present when the sounds were played (ostensibly causing the sounds, although the idea of causality was not explicitly stated in the instructions). Across conditions, we manipulated the timing of the musical sounds such that one of two *sound sequences* played: a descending scale with evenly-timed notes or unevenly-timed notes. The two sound sequences were timed such

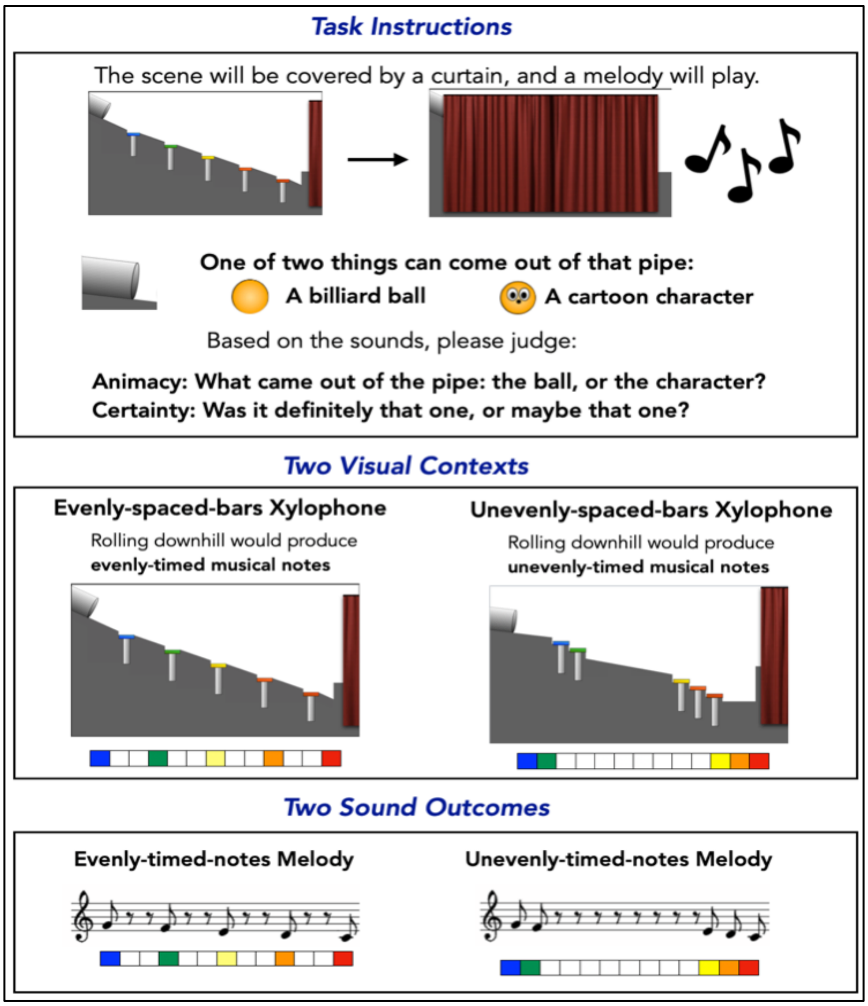

Figure 2.1: Experiment 1a Method. Participants were asked to judge whether an animate agent or an inanimate object was behind a curtain when musical sounds occurred. We manipulated the visual context and the timing of the musical sounds that occurred to test whether people integrated auditory and visual information to infer whether an agent caused the sounds.

that in the evenly-spaced visual context, a ball rolling downhill due to gravity would be expected to play the evenly-timed sound sequence; in the unevenly-spaced visual context, the ball would be expected to play the unevenly-timed sound sequence.

If people can use causal reasoning to infer that animate agents cause musical sounds, then people should make different inferences about whether an agent was the cause of each of these sound sequences depending on the visual context in which it was produced. That is, there should be an interaction between the nature of the sound and its visual context. In a visual context where a sound sequence could be produced by gravity alone, the possible physical-mechanical explanation should "explain away" the agent, making people less likely to infer that an agent caused the sounds. In contrast, in a context where the sound sequence seems to require self-propelled movement (e.g., jumping and changing speed or direction), people should infer that the sound was caused by an agent, thus changing their inferences about the cause of the sound sequences depending on their context.

Alternatively, people may believe that all musical sounds are created by an agent (Launay, 2015). In this case, participants should answer that all of the sound sequences were more likely to be produced by the agent, regardless of visual context. Similarly, participants may expect that musical sounds with certain features (e.g., more orderly sequences, Newman et al., 2010) are always associated with agents. In this case, participants should infer that the evenly-timed notes were produced by an animate agent equally in both visual contexts. These patterns of findings would suggest failure to engage in causal reasoning and failure to consider alternative physical-mechanical explanations for the sounds that were suggested by the visual context. The study design, sample size, and analyses, described below, were pre-registered before data collection (https://aspredicted.org/rf47e.pdf).

Methods

Participants. As indicated in the pre-registration, sixty undergraduate students were recruited from an online study pool at a large public university in Southern California (Age: M = 20.97 years, Range = 18 to 31; 19 males) and tested online. All participants gave informed consent and received credit as compensation. Seventeen additional participants were tested but ultimately excluded due to the pre-registered exclusion criteria: failing one or more attention check questions (13) and technical issues (4).

Stimuli. Animated videos and accompanying sounds were constructed using Apple Keynote, Apple GarageBand, and Apple QuickTime software. Two familiarization videos demonstrated how the xylophone-like bars produced sounds when struck (one video for each visual context). Both videos showed a hill with colored bars and a pipe at the top of the hill (see Figure 2.1). A mallet entered the scene and struck a sequence of bars before exiting the scene. The sequences played in the familiarization videos were designed to differ equally from both sound sequences played during the test trials, and thus they consisted of the ascending scale played twice, once with even timing and once with uneven timing. Two additional familiarization videos provided evidence of the ball's inanimacy and the character's animacy by showing the character moving in a self-propelled way and the ball rolling with gravity.

During each trial, a test video was shown in which the visual context was covered by an animated curtain, and one of two sound outcomes was played: a descending scale with evenly-timed notes (G, F, E, D, C, with fundamental frequencies of 778, 702, 647, 569, and 511 Hz), or a descending scale with unevenly-timed notes (the same pitches, played G-F-E-pause-pause-D-

C). Both sound sequences were of equal duration overall (4 seconds).

Design. The experiment used a 2 (visual context) x 2 (sound outcome) within-subject design. Each participant completed four trials composed of two blocks (by visual context) of two trials each (one for each sound outcome). The order of the blocks and the order of test trials within a block were counterbalanced between subjects.

Procedure. Participants first viewed a familiarization video demonstrating how the bars produced sounds when struck. Then, they were asked to describe what happened and answered two attention check questions (whether each bar played the same note, different notes, animal sounds, or no sounds and whether the xylophone bars were evenly-spaced or unevenly-spaced).

Participants were then asked to notice the pipe on the left-hand side of the scene and told that one of two things could come out of the pipe: a ball or a cartoon character. To provide evidence of the ball's inanimacy and the character's animacy, brief videos were shown (of the character moving in a self-propelled way and of the ball rolling with gravity), and they were identified in ways consistent with inanimacy and animacy, specifically as a billiard ball or as cartoon character with a name (Fred).

Participants then completed test trials. During each test trial, participants watched a video where the visual context was occluded and a sound sequence was played (either the evenly-timed or unevenly-timed descending scale). Participants were then asked to judge whether an animate agent or an inanimate object had been present in the scene when the sounds occurred using a two-alternative forced choice ("What came out of the pipe? Was it the ball? Or was it Fred?"). Participants then reported their certainty about this answer, again using a two-alternative forced choice ("Was it definitely that one, or maybe that one?"). This response format was used in order to allow for developmental comparison between adults and children (see Exp. 1b). The two scales were then combined to create a single four-point scale (1 = Definitely the ball, 2 = Maybe the ball, 3 = Maybe the agent, and 4 = Definitely the agent).

After completing two test trials in the first block, participants received similar instructions for the second block (the other visual context condition) with a new visual context introduced. They then completed two additional test trials similar to those in the first block, but with videos showing the new visual context. Lastly, participants were asked if they experienced any technical issues and to guess what the experiment was about before submitting their answers.

Results

To test whether participants integrated auditory information with visual information to infer the causes of musical sounds, we used an ordinal logistic regression model to predict participants' judgements of what caused the sound (Definitely the ball = 1 to Definitely the agent = 4; ordinal variable) from the predictors of the visual context (evenly-spaced-bars xylophone or unevenly-spaced-bars xylophone), the sound outcome (evenly-timed notes or unevenly-timed notes), the interaction of these two factors, and participant (as a random effect). As predicted, there was a significant interaction between visual context and sound outcome ($\chi^2(1) = 112.21$, $p < 0.001$, nested model comparison; $Z = - 9.51$, $p < 0.001$, coefficient test; see Figure 2.2). In the context of the evenly-spaced-bars xylophone, participants judged it more likely that the agent had been present after hearing the unevenly-timed notes ($M = 3.57$, $SE = 0.10$) than the evenly-timed-notes ($M = 1.72$, $SE = 0.10$; Wilcoxon Test, $Z = -6.13$, $p < 0.001$). This pattern was reversed in the other visual context. When presented with the unevenly-spaced-bars xylophone, participants judged it more likely that the agent had been present after hearing the evenly-timed notes ($M = 2.94$, $SE = 0.14$) than the unevenly-timed notes ($M = 1.92$, $SE = 0.11$; Wilcoxon Test, $Z = -3.98$, $p < 0.001$). There were also significant main effects of the visual context ($Z = 6.44$, $p < 0.001$) and of the sound outcome ($Z = 8.90$, $p < 0.001$, see Figure 2.2).

Experiment 1a provides evidence that adults can use causal reasoning to link musical sounds with animate agents. When the timing of notes required changing speed or direction, adults inferred that they were caused by an animate agent. In the context of an evenly-spaced-bars xylophone, adults more often inferred that an agent caused the unevenly-timed notes, as producing the sounds required self-propelled movements beyond rolling down by gravity. This pattern was flipped in the context of the other xylophone: When the bars were spaced unevenly, adults more often inferred that an agent caused the evenly-timed notes, and adults chose the agent less often for the unevenly-timed notes. This supports the idea that adults were able to

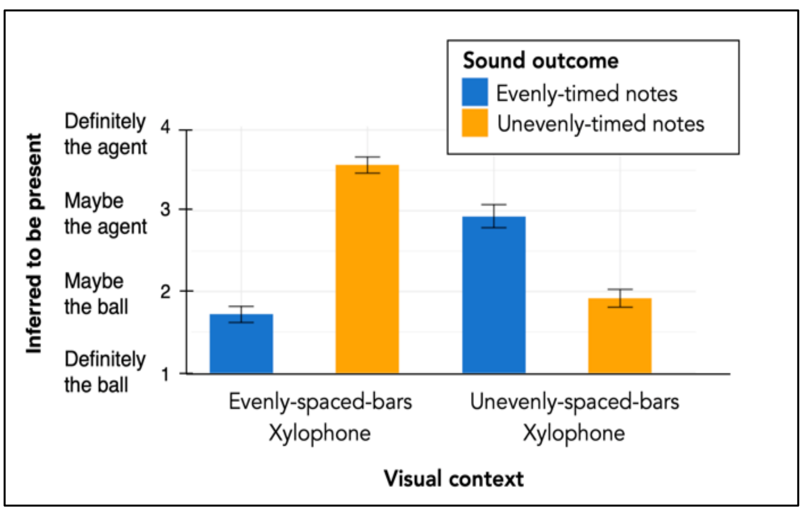

Figure 2.2: Experiment 1a Results. Participants' judgments of whether an animate agent (the character) or an inanimate object (the ball) had been present when the sounds played for each visual context (evenly-spaced-bars xylophone, unevenly-spaced-bars xylophone) and sound outcome (evenly-timed notes, unevenly-timed notes).

integrate auditory information from the musical sounds with visual information about the environment, and use causal reasoning to infer which musical sounds must be caused by self-propelled movements and animate agents. In Experiment 1b, we investigate how this ability to flexibly reason about sounds develops in childhood. Do children also show adult-like reasoning about what musical sounds were caused by animate agents vs inanimate objects? Or, are children more likely to use surface features or focus on only one modality (e.g. properties of the sound), failing to integrate auditory and visual information? We use a similar experiment, with an experimental design appropriate for 4- to 6-year-olds, to examine whether children can also engage in causal reasoning about the causes of musical sounds.

Experiment 1b

We know that links between music and movement emerge early in life (e.g., Kim & Schachner, 2022; Kragness et al., 2022). Do young children, then, use causal reasoning to link musical sound with movement and animacy? One possibility is that this type of reasoning may be adult-like in children from early life. Infants have early-developing intuitive theories of the physical world and social world (e.g., Spelke & Kinzler, 2007). This could allow very young children to reason about the causes of musical sounds.

Alternatively, this reasoning may undergo developmental change during the pre-school years. Previous studies have shown that infants and young children exhibit a bias in expecting that orderly things must be created by animate agents (Newman et al., 2010; 2015; Ma & Xu, 2013). Younger children may always link orderly musical patterns with animate agents. Or, young children may think all musical sounds are caused by animate agents (Launay, 2015).

Lastly, children below age 6 could differ due to failures to integrate visual with auditory information to reason about the causes of sounds. This would be in line with a recent developmental study of causal reasoning about non-musical sounds, in which children before age 6 did not engage in causal reasoning about the movements required to cause the sounds, and did not integrate vision and sound (Outa, Zhou, Gweon, Gerstenberg, 2022). Therefore, we tested whether and how this reasoning occurred with 60 4- to 6-year-old children online over the Zoom platform.

Methods

Participants. As indicated in the pre-registration (SRCD 2021), N=60 children participated, including twenty four-year-olds (Mage = 54.1 months, 11 females), twenty five-year-olds (Mage = 66.3 months, 9 females), and twenty six-year-olds (Mage = 77.6 months, 5 females). Participants were recruited through the lab online database of local families interested in research, and participated in a 30-minute video call session via the Zoom videochat platform. All participating families gave informed consent and received a $5 online gift card as their reward. Three additional participants were tested but ultimately excluded due to the pre-registered exclusion criteria: failing to finish some trials (1) and technical issues (2).

Design and Stimuli. The stimuli videos and the 2 x 2 design of musical sounds x physical environments were identical to those of Experiment 1a with three minor changes. First, to enable a synchronous interactive research game with children, all stimuli were embedded in a slide show from an online slideshare platform (slides.com), and parents were asked to open a link to a slide show. Parents navigated to the slideshow directly in their own web browser to avoid audio-visual delays in stimulus presentation. The experimenter remotely controlled the slides on the slideshow while also interacting with the child over the Zoom platform.

Second, the instruction that during the task a curtain would come in and cover the scene, which was presented as text for adults, was also visually-depicted for children using a short video of the curtain coming in and covering the xylophone during the instruction block. Third, during the test trials, a picture of agent (Fred) and the ball were added to the slides so that children could opt either to verbally answer, or to point to the pictures to answer the main test question (e.g., "What came out of the pipe, Fred or the ball?"). The pictures of the agent and object appeared on each side of the bottom of the slides, after the test video ended (location of the agent and the ball was counterbalanced across participants). To practice with this interface and provide a calibration for the experimenter, before the test trials children were asked to practice pointing to the agent and the ball with their finger.

Procedure. The study procedure in Experiment 1b aligned with that of Experiment 1a, with minor changes to make the task appropriate for 4- to 6-year-old children. All questions and instructions were verbally spoken by the experimenter over videochat (Zoom). Participants could respond by pointing to the side of the screen with their finger (e.g., left or right, counterbalanced) or verbally respond using words (e.g., Fred or the ball, see Figure 2.3).

At the beginning of the videocall on the Zoom platform, the experimenter greeted the parent and the child participant and checked whether the audio and video were working on both ends. Then, the experimenter confirmed the parent's consent for video recording and shared a link to the slideshow for displaying the stimuli slides (slides.com). When the parent opened the link, the experimenter explained that the slides will be remotely controlled by the experimenter.

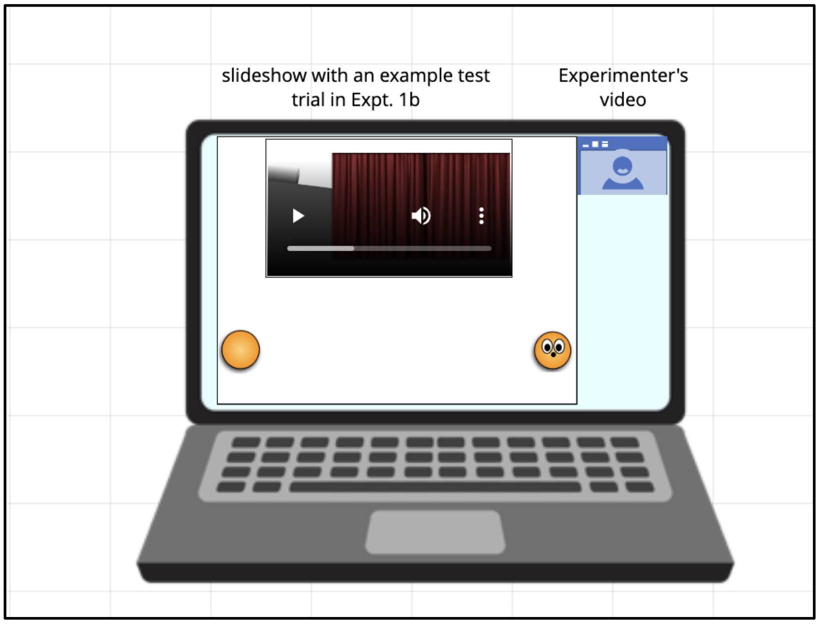

Figure 2.3: An example set-up for a test trial in Experiment 1b. In each test trial, the stimuli video was played and then the pictures of the two potential causes (the ball and Fred) appeared on the screen. Children were asked to answer what came out of the pipe either verbally or by pointing to one of the two pictures (e.g., picture of Fred). Then, the experimenter asked the child how sure they were about their answer (e.g., Maybe Fred, or definitely Fred). Parents were asked to sit next to the child and help clarify their answers in case children's pointing behavior was not fully captured on the camera.

The experimenter then guided the parents on setting up the screen to see both the slideshow and the experimenter's face, but not the child's face themselves to minimize distractions. Parents were asked to hide their self-view, make the experimenter's video smaller and move it to the top corner of the screen, and expand the web browser with the stimuli as big as possible for the rest of the screen. Lastly, to check the audio setting on the slideshow, the participant watched a short animation and confirmed that they could hear the video.

As Experiment 1a, child participants in Experiment 1b were first introduced to one of the two physical contexts (e.g., the xylophone staircase with the bars evenly or unevenly spaced). The xylophone was first introduced with a picture where the experimenter guided the child's attention to the spacing of the musical bars of the first xylophone with the mouse cursor (e.g., for the xylophone with evenly-spaced bars, “The bars are spread out evenly across the hill”; for the xylophone with unevenly-spaced bars, “The bars are all spaced funny, see there are some up here, some down here, and none in the middle!”). Then children watched a familiarization video showing how the xylophone works. After watching the video, the experimenter asked the child to notice the pipe on the left-hand side of the scene, and were told that one of two things could come out of the pipe: An inanimate ball, or an animate character (Fred). Then, in the following two slides, children watched a brief video of a ball rolling up and down on a hill by gravity with some descriptions (e.g., “This ball is really heavy and really hard! Here’s a little video of the ball”) and a brief video of the animate character (Fred) jumping and moving back and forth in the video (e.g., “Let’s meet my friend Fred! Here he is! Hi Fred!”).

After learning about the first xylophone and the ball and agent, participants were instructed that they will be playing a game where they try to figure out what came out of the pipe but that the scene would be covered by a curtain. The participant watched a video of the curtain coming in and covering up both the xylophone and the end of the pipe along with the following description: “We are going to cover everything up with a curtain, so you can’t see! But you will get to hear some sounds from the xylophone! You can use the sounds to make a guess about what came out of the pipe.” Then, the two questions on animacy judgement and certainty of their answers were introduced and practiced. The experimenter then told them that they will be asked to guess what came out of the pipe. The experimenter animated in the pictures of the ball and Fred on each side of the screen (e.g., left/right, counterbalanced across participants) one by one and participants were guided to practice pointing to each of them with their finger. Since the zoom platform defaults to mirroring the videos, the experimenter took note of which side of the screen the child was pointing to for each item (e.g., when the child points to the picture on their right side , the pointing finger would appear on the experimenter's left side). Then, the experimenter guided the child on how to report how sure they were about their answers by using 'definitely' when they are sure and 'maybe' when they are not sure.

The test trials were blocked, such that the two trials with the same xylophone were presented back-to-back. Across participants, the order of xylophones and the order of sound sequences were counterbalanced.

In each test trial, a video was played where the visual context (the xylophone) was occluded, and one of the two sound sequences were played (evenly-timed-notes melody, unevenly-timed-notes melody; order counterbalanced across participants). After watching. the child was asked to answer a two-alternative-forced-choice question as the experimenter displayed the pictures of the ball and Fred: "What came out of the pipe? Was it the ball? Or was it Fred?" The order of the ball and Fred were counterbalanced across the participants. When the child said their answer or pointed to one of the two pictures, the experimenter asked "Was it definitely (the ball/Fred) or maybe (the ball/Fred)?". In analysis, these two scales were combined, to create a single four-point scale (Definitely the ball, Maybe the ball, Maybe the agent, Definitely the agent).

Results

We first analyzed the full sample using the same ordinal logistic regression model used for adults, but with age as an added factor, as well as the 3-way interaction between age, sound outcome, and visual context. This model thus predicted participants' judgements of what caused the sound (Definitely the ball = 1 to Definitely the agent = 4; ordinal variable) from the predictors of the visual context (evenly-spaced-bars xylophone or unevenly-spaced-bars xylophone), sound outcome (evenly-timed notes or unevenly-timed notes), the interaction of these two factors, age (in months, as a continuous variable), the interaction of these three factors, and participant (as a random effect). We found a significant 3-way interaction of age in months*sound outcome*visual context, meaning that the way that sound outcome interacted with visual context changed with age (β = -0.15, p = 0.002). In addition, we found three significant 2-way interactions (age in months*sound outcome, β = 0.12, p = 0.0006; sound outcome*visual context, β = 0.53, p = 0.003; age in months*visual context, β= 0.09, p = 0.007), and significant main effects of all three factors (sound outcome, β = -8.29, p = 0.0002; visual context, β= -5.76, p = 0.009; age in months, β= -0.05, p = 0.04).

To understand differences by age in children's judgements, we next separately examined data within each age bin. We found that like adults, 6-year-old children showed a significant 2 x 2 interaction of visual context and sound outcome, in line with the use of causal reasoning. β = -2.53, p = 0.003; also confirmed through a nested model comparison, *X^2*(1) = 9.12, p = 0.003). In particular, they thought the sounds were caused by an agent for different musical sounds depending on the context. In the context of the xylophone with unevenly-spaced bars, the agent was judged to be present more often for evenly-timed notes (*M* = 3.10, *SE* = 0.20) than unevenly-timed notes (*M* = 2.20, *SE* = 0.21; Wilcoxon Test, *Z* = 53.0, p = 0.26). In contrast, in the context of the xylophone with evenly-spaced bars, the agent was judged to be present more often for the unevenly-timed notes melody (*M* = 2.65, *SE* = 0.21) than the evenly-timed notes melody (*M* = 2.25, *SE* = 0.24; Wilcoxon Test, *Z* = 122.5, p = 0.03). A significant difference between the two sound outcomes was observed in the unevenly-spaced xylophone but not in the evenly-spaced xylophone (Wilcoxon Test, Z = 122.5, p = 0.03).

In contrast, in 4-year-olds, we found a very different pattern. An ordinal logistic regression model did not reveal a significant 2x2 interaction of the visual context and the sound outcome (β = 0.80, p = 0.33). Instead, we found a significant main effect of sound outcome (β = -1.64, p = 0.007; also confirmed by nested model comparison, $X^2(1)$ = 8.97, p = 0.003). In both contexts, 4 year old children judged it significantly more likely that the agent had been present after hearing the evenly-timed notes melody than the unevenly-timed notes melody (xylophone with evenly-spaced bars, evenly-timed notes: M = 2.95, SE = 0.23, unevenly-timed notes M = 1.90, SE =0.26; Wilcoxon Test, Z = 135.0, p =0.03; Xylophone with unevenly spaced bars, evenly-timed notes M = 2.60, SE = 0.28, unevenly-timed notes, M = 2.10, SE = 0.27, Wilcoxon Test, Z = 105.0, p = 0.17). There were no other significant effects.

Lastly, 5-year-olds did not show a 2x2 significant interaction (visual context*sound outcome, β = 0.87, p = 0.34) nor significant main effects of the two variables (visual context, β = -0.33, p = 0.55; sound outcome, β = -0.76, p = 0.18).

Discussion

We found evidence that 6-year-old children can use causal reasoning to link music with animate agents, showing judgements similar to adults'. This also shows that children at this age can integrate auditory and visual information in their causal reasoning. In contrast, 4 year old children showed a different pattern, judging that the evenly-timed notes were more likely caused by an agent across both xylophone contexts. Five-year-olds in our sample were somewhere in the

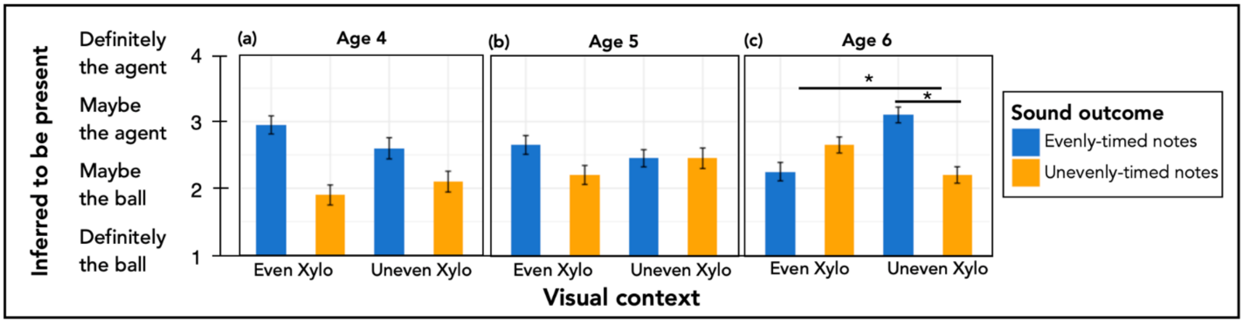

Figure 2.4: Results in children age 4-6 years in Experiment 1b. Different inferences on the cause of music by the sound outcome and visual context in age 4,5 and 6. Panel (a), (b) and (c) shows results from 4-year-olds (n = 20), 5-year-olds (n = 20), and 6-year-olds (n = 20), respectively. The y-axis shows children's judgments on whether an animate agent (Fred) or an inanimate ball had been present when the sounds were played. The x-axis shows the two visual contexts (xylophones with evenly-spaced-bars vs. unevenly spaced bars) for each age group. The sound outcome is color-coded such that the blue bars indicate the ratings for evenly-timed notes melody and the orange bars show the ratings for the unevenly-timed notes melody. Error bars indicate the standard errors.

middle, showing at-chance responding that was potentially a mid-point between the patterns seen in 4- and 6-year-olds. At this age, children may be shifting from one type of reasoning to another, or show individual differences potentially due to differences in cognitive development across children.

These data provide evidence that from 6 years of age, high-level cognition can link music with social concepts through rational inferences about the movements and agents that cause the sounds. Interestingly, while six-year-olds seemed to be reasoning about what had to happen behind the curtain to cause the sounds, 4-year-olds were inclined to map orderly sounds (e.g., evenly-timed notes melody) to the agent (vs. the ball). This is consistent with the idea that these young children have a bias toward linking orderly sounds with animate agents (Newman et al., 2010; Ma et al., 2013), and may not fully integrate auditory and visual information in their reasoning, in line with previous work (Outa et al., 2022). This suggests developmental change, in that younger children use perceptual features of music to make judgements, while older children engage in causal reasoning about musical sounds that integrate both auditory and visual information.

Experiment 2a

Our first experiments provide evidence that adults and school-age children can use causal reasoning to infer when an agent caused a musical sound. In the next experiments, we ask: How flexible is the system of reasoning that supports these inferences?

People may have a rich and flexible causal model of actions and the environment. If so this would involve representing the multiple factors that interact to cause observed sounds (such as the musical sounds we explore here). A plausible generative model of this is depicted in the accompanying Bayes Net (Figure 2.5). This model can be summarized as follows: Sounds are caused by movements. The movements are in part caused by the structure of the environment (Is there a hill? Is there a round object on top of the hill?), as well as the physical inanimate forces that are present (gravity, wind). For movements caused by agents, however, the path of the movements are jointly caused by another factor as well: The agents' mental states (goals, desires and beliefs). This framework helps to explain how people inferred the presence of an agent in Exp 1: When the movements could not be well-explained by physical forces and the structure of the environment, another explanation was invoked: The movements were impacted by another force, the intentional self-propelled actions of an agent.

If people are representing the causal system in the way depicted in Figure 2.5 this would allow them to observe any subset of variables and infer any remaining latent variable. Thus, they should first be able to observe the sounds and environment, and from this infer the movements and presence of agents (as shown in Exp. 1). In addition, people should be able to invert the system to infer the nature of the environment. If people observe whether an agent is present, or not, this should constraint what types of movements would be possible (e.g. self-propelled movement). Then, after observing what sounds were produced, people should make different inferences about the nature of the environment. If no agents are present, then the environment must be structured such that the inanimate objects and forces could plausibly cause the sounds (e.g. see Methods Figure 2.6). If an agent is present, the nature of the environment should be less certain (vs. when no agent was present), as the agent's ability to change speed and direction could allow it to create the sounds in many different environments.

In Experiment 2a, we test the hypothesis that people can also invert this inference to infer the structure of the physical environment based on hearing the sound and knowing whether an animate agent or inanimate object was present in the scene.

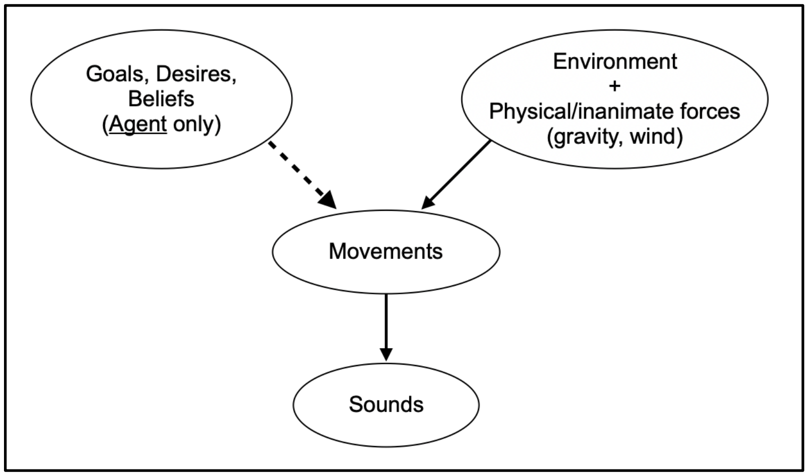

Figure 2.5: A generative model of sound production.

Methods

Participants. Fifty-four undergraduate students (Mean age = 20.8 years, SD = 3.14, 40 females) at a public university in Southern California participated in an online study and received course credit. Six additional students participated, but were dropped due to failing one or more comprehension check questions (6).

Design. The experiment used a within-subject 2x2 design with 4 trials per participant. The independent variables were the *sound outcome* (evenly-timed notes or unevenly-timed notes) and whether the animate agent or the inanimate object was present. The dependent variable was participants' judgment of what physical environment was behind the curtain, the evenly-spaced xylophone or the unevenly-spaced xylophone (see Figure 2.6).

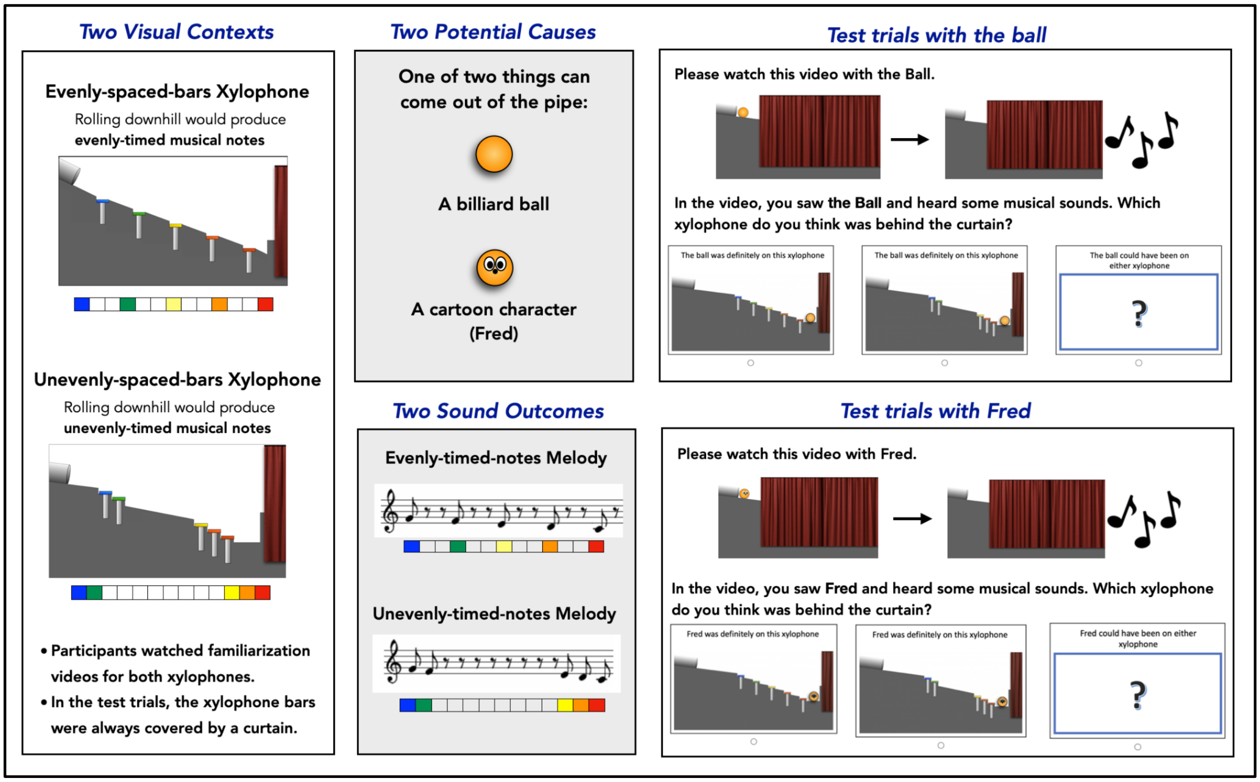

Figure 2.6: Experiment 2 Methods. In Experiment 1, the physical environment and sound outcome were provided for the participants to judge the agent (the cause of the music). In Experiment 2, the potential causes and the sound outcome were provided for the participants to judge which xylophone was behind the curtain (the evenly-spaced xylophone, unevenly-spaced xylophone, or either xylophone).

Procedure and stimuli. As in Experiment 1, participants were familiarized with the two musical causes (the ball and the agent, Fred) and the two types of physical contexts (the evenly-spaced xylophone and the unevenly-spaced xylophone). Then, participants were instructed that they would see a curtain covering the xylophone and that one of the two causes would come out of the pipe, the ball or Fred.

The four test trials were blocked in pairs, by sound sequence (either the evenly-timed notes or the unevenly-timed notes), such that participants would hear the same sound sequence for the first two trials and then the other sound sequence for the last two trials (counterbalanced across participants), alternating between the cause of the music (the ball or Fred; order counterbalanced across participants).

During each trial, a video played in which the xylophone was already covered by a curtain at the start of the video. Then, either the object or agent came out of a pipe located at the top of the hill, and moved down the hill toward and behind the curtain. After the object or agent was behind the curtain, one of two sound outcomes played: evenly-timed notes or unevenly-timed-notes. Participants were then asked to answer a multiple-choice question: “Which xylophone was behind the curtain?” Three answer choices were presented as pictures: (1) a picture of the evenly-spaced xylophone; (2) a picture of the unevenly-spaced xylophone; and (3) a box with text stating “it could have been either one” (i.e., either xylophone behind the curtain). Participants were then asked to explain why they chose their answer (“What made you think that?”) as a free-response question.

Results

Preliminary analysis showed that there were no significant effects of the order of the physical context (left-right positions of the two xylophones in test trials, which matched the order of the familiarization videos of the two xylophone contexts, the order of the two potential causes, or the order of the sound outcome blocks presented on the dependent measure of judgment of the physical context behind the curtain (order of the physical context, $\chi^2(1) = 0.92$, $p = 0.66$; order of the sound outcome, $\chi^2(1) = 2.86$, $p = 0.22$; order of the causes, $\chi^2(1) = 2.11$, $p = 0.34$, chi-square test of independence).

First, we tested the prediction that participants should have more uncertainty about the physical environment when the agent was present in the scene, vs. when the ball was present in the scene, and should therefore choose “it could be either one” more often in the agent vs. ball conditions. To test this, we used a logistic regression, in which the dependent variable (DV) was whether they chose “it could be either one”, or not (thus re-coding the data as a binary variable). The predictors were the sound sequence (evenly-timed notes and unevenly-timed note), the two causes (ball, agent), the interaction between these two factors, and participant (as a random factor). The multiple regression model showed a significant main effect of the sound outcome (β = -6.79, p = 0.0002), and no other significant effects. However, in a targeted comparison testing our prediction, we found that participants chose “could be either one” more often (regarding which xylophone was behind the curtain) when the agent was present than when the ball was present (20.3% vs. 3.7%, $\chi^2(1) = 52.07$, $p < 0.0001$, McNemar’s test). This was true whether the evenly-timed notes were played (16.7% vs. 3.9%, $\chi^2(1) = 28.9$, $p < 0.0001$, McNemar’s test), or whether the unevenly timed notes were played (24.1% vs. 3.7%, $\chi^2(1) = 22.2$, $p < 0.0001$, McNemar’s test). We also note that only the minority of answers were “it could be either one” (12.0% overall), with participants choosing one of the xylophones on the majority of trials (88.0% overall).

Second, we examined participants’ other answers (non-“either one” choices), using multiple logistic regression with the DV whether they chose the evenly-spaced xylophone, or the unevenly-spaced xylophone. The predictors were the same: the sound sequence (evenly-timed notes and unevenly-timed note), the two causes (ball, agent), the interaction between these two factors, and participant (as a random factor). We found a significant effect of melody type (β = -

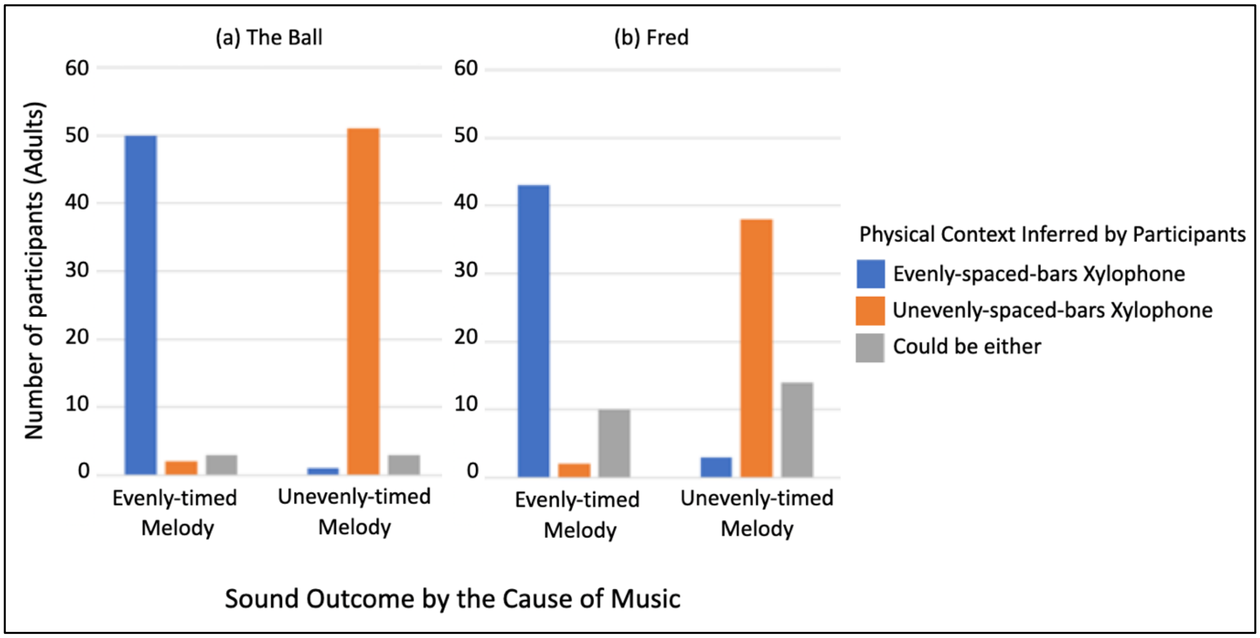

Figure 2.7: Results for adult participants in Experiment 2a. Panel (a) shows the ratings for the ball and panel (b) shows the ratings for Fred as the cause of music. Number of adult participants in Experiment 2a that chose each type of possible physical context (evenly-spaced xylophone, unevenly-spaced xylophone, or could be either one) by the cause of the sounds (Fred or the ball) and the type of sound outcome (evenly-timed notes or unevenly-timed notes).

5.90, $p < 0.001$) and no other significant effects. Participants were more likely to pick the evenly-spaced xylophone after hearing the evenly-timed notes than the unevenly-timed notes, both in the ball condition (96.1% vs. 2.0%) and in the agent condition (95.6% vs. 7.3%).

Discussion

Experiment 2a showed that adults can flexibly infer the physical environment (evenly-spaced-bars xylophone vs. unevenly-spaced-bars xylophone vs. could be either) that must have been present by reasoning from what the cause of music was (an inanimate object vs. an animate agent) and what melody was played (evenly-timed-notes melody vs. unevenly-timed-notes melody). Adults were more likely to reason that if an inanimate object was playing a melody on the xylophone, the spacing of the xylophone bars in the physical context should match the timing of the notes played in the melody. Similar patterns were observed in the reverse direction with the animate agent (Fred), that adults inferred that a melody that matches the spacing of the bars must have been played although there were numerically more responses that chose ‘either’ for the animate object as the cause of music.

Interestingly, not all adult participants chose “either” xylophone as the physical context when the animate agent was present and instead showed a bias toward mapping the characteristics of the sound outcome (e.g., the timing of the musical notes within the melody) to the constraints in the physical context (e.g., the spacing of the xylophone bars). Why do many people answer that the environment matches the sound, even in the agent condition? This may be because the matching sound pattern is the easiest to produce (would require the least effort from the agent). People expect other agents to act in ways that minimize cost (such as effort) while maximizing reward (termed the naive utility calculus, e.g. Jara-Ettinger et al 2015). Since participants do not have any reason to expect that one sequence of tones would be strongly preferred over the other, they may reason that the agent is likely to take the easiest path downhill, rather than making additional effort to change speed and direction to produce the non-matching melody.

Still, adults could perform flexible reasoning such that they represented the animate agent as having more capacity to overcome the physical constraints posed by the xylophone, perhaps by jumping between different bars. Therefore, we used a similar design in Experiment 2b to test whether 4-year-olds and 6-year-olds demonstrated similar patterns of flexible causal reasoning about sounds as those observed in Experiment 1b.

Experiment 2b

In Experiment 2b, 4- and 6-year-old children were presented with the stimuli from Experiment 2a. These two age ranges were tested because, in Experiment 1b, they showed different patterns of judgements when presented with a similar task: 6-year-old children showed causal reasoning similar to adults', whereas 4-year-old children showed surface-level mapping of the more orderly sound outcome (evenly-timed notes) to the animate agent. Since 5-year-olds in Experiment 1b did not show a statistically meaningful pattern, only the performance of 4- and 6-year-old children was compared in Experiment 2b.

We predicted that 6-year-olds can infer the structure of the physical environment based on hearing the sound, and knowing whether an animate agent or inanimate object was present in the scene. We predict that 4-year-olds may fail this task based on our results from Experiment 1a and we aim to examine whether they still show a similar bias – linking any orderly sounds to the action of an agent regardless of the visual context.

Methods

Participants. N=39 children, 19 4-year-old and 20 6-year-old children participated in an online study conducted on a video call platform (Zoom). Participants were recruited from online advertisements on the lab's webpage and also from a database of local families interested in participating in research. An additional 5 children participated in the study but were excluded due to failing comprehension check questions (4) and having a wrong birth date provided (1).

Design and Stimuli. The 2 x 2 design and the stimuli videos were identical to those of Experiment 1a.

Procedure. The study procedure in Experiment 2b was identical to that of Experiment 2a, with minor changes to make the task appropriate for 4- and 6-year-old children, analogous to the study design in Experiment 1b. An instructional block where children could practice pointing to the three answer choices (evenly-spaced xylophone, unevenly-spaced xylophone, or either) was added. Also, after each test trial, a comprehension check question asked the child whether the animate agent or the inanimate object had been present in the video to make sure that participants were paying attention to the stimuli (e.g., "Which was in the video? Fred? Or the ball?").

Results

Preliminary analysis showed that there was no significant effect of the order of the xylophones presented ($\chi^2(1) = 3.52$, $p = 0.17$), the order of the sound outcomes presented ($\chi^2(1) = 4.83$, $p = 0.08$) or the order of the two causes (the ball, Fred) presented ($\chi^2(1) = 2.08$, $p = 0.22$) on our dependent variable (choice of the visual context). Therefore, these variables were merged together for the full data analysis.

First, we tested the prediction that participants should have more uncertainty about the physical environment when the agent was present in the scene, vs. when the ball was present in the scene, and should therefore choose "it could be either one" more often in the agent vs. ball conditions. To test this, we used a logistic regression, in which the dependent variable was whether they chose "it could be either one", or not (choosing only one xylophone among ones with evenly-spaced bars or unevenly-spaced bars). The predictors were the sound sequence (evenly-timed notes and unevenly-timed note), the two causes (ball, agent), the interaction between these two factors, age group (4 vs. 6 years old), the interaction between age group and the two causes (ball vs. agent), the interaction between age group and the two sound sequences, the three-way interaction between all three factors, and participant (as a random factor). In this multiple logistic regression, we found a significant main effect of age group (β = 1.48, p = 0.023). The 4-year-old children were more likely to choose "could be either one" than the 6-year-old children (35.5% vs 13.8% overall, $\chi^2(1) = 22.82$, $p < 0.0001$, McNemar's test). There was no significant effect of ball vs. agent, unlike adults' judgements ($\beta = 1.73$, $p = 0.55$). There was also no significant interaction of age with ball vs. agent, as would be expected if one age group but not the other showed adult-like patterns of reasoning ($\beta = -0.32$, $p = 0.60$).

Second, we examined participants' other answers (non-"either one" choices), also using a multiple logistic regression. Now the DV was whether they chose the evenly-spaced xylophone, or the unevenly-spaced xylophone. The predictors were the same: the sound sequence (evenly-timed notes and unevenly-timed note), the two causes (ball, agent), the interaction between these two factors, age group (4 vs. 6 years old), the interaction between age group and the two causes (ball vs. agent), the interaction between age group and the two sound sequences, and participant (as a random factor). We found a significant interaction between age group and sound sequence ($\beta = -2.06$, $p = 0.005$), a significant main effect of age group ($\beta = 1.35$, p= 0.04), and a significant main effect of the sound outcome ($\beta = 9.99$, $p = 0.007$). Six-year-olds but not four-year-olds were more likely to pick the xylophone with evenly-spaced bars (vs. unevenly-spaced bars) after hearing the evenly-timed notes than the unevenly-timed notes (6-year-olds: 43.5% vs 4.3%, $\chi^2(1) = 37.8$, $p < 0.0001$, chi-square test of independence; 4-year-olds: 28.6% vs 18.4%, $\chi^2(1) = 1.02$, $p = 0.31$, chi-square test of independence). These results did not differ by the causes (ball or Fred) in both age groups (See Figure 2.8). There were no other significant effects.

Did 4-year-olds respond differently to the three choice options (Evenly-spaced-bars xylophone, Unevenly-spaced-bars xylophone, Either) from what the 6-year-olds did when merged across different conditions? A confirmatory analysis on the choice patterns of the two age groups (4-year-olds, 6-year-olds) was conducted with the chi-square test of homogeneity. Results confirmed that both samples were heterogenous ($\chi^2(2) = 10.04$, $p = 0.007$, chi-square test of homogeneity), meaning that the two age groups were heterogenous and showed statistically significant differences in their distributions between the 3 levels of the dependent variable. Did the 4-year-olds in our dataset show chance performance at all three levels of the dependent variable (Evenly-spaced-bars xylophone, Unevenly-spaced-bars xylophone, Either)?

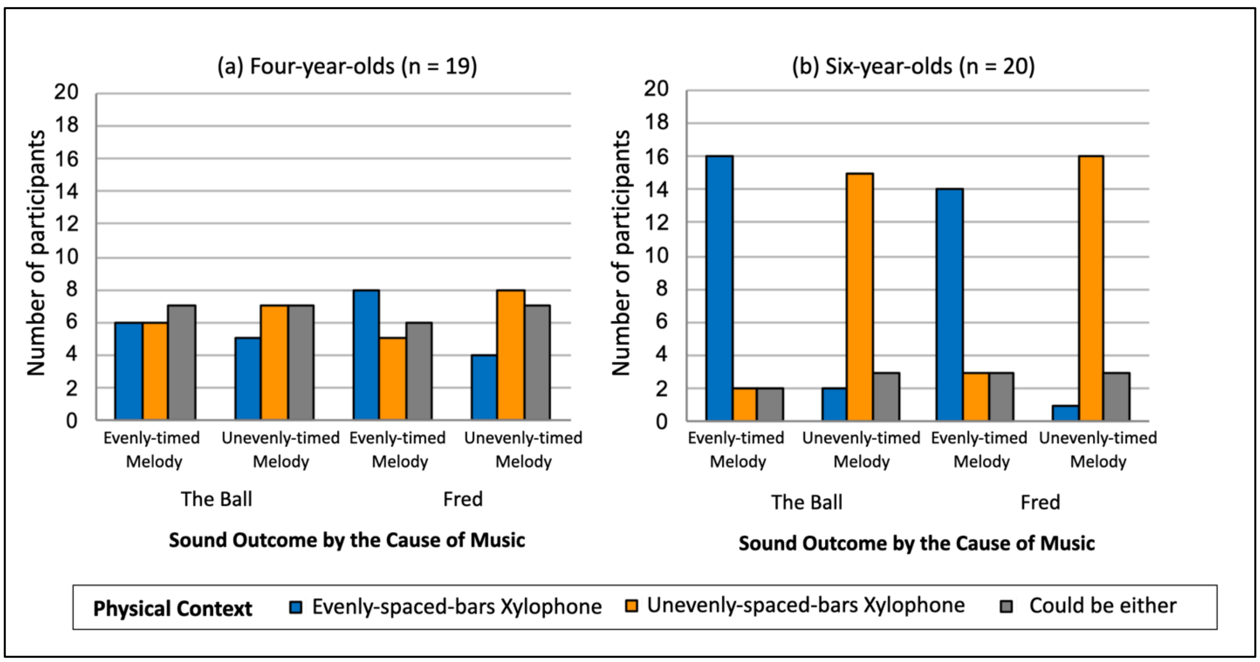

Figure 2.8: Number of participants choosing each physical context option by sound outcome and cause of music in Experiment 2b. Panel (a) shows the responses from the 4-year-olds (n = 19) and panel (b) shows the responses from the 6-year-olds (n = 20). The x-axis shows the four test trial types (Sound outcome x Cause of music) and the y-axis shows the number of responses for the three possible choices of the physical context inferred by our participants.

A chi-square test of goodness of fit compared the proportion of each response to baseline probability (33.3%). Results indicated that the 4-year-olds were not significantly different from this baseline level ($\chi^2(2) = 0.34$, $p = 0.84$, Chi-square test for goodness of fit). In contrast, the observed proportions of responses in 6-year-olds in the three answer choices (Evenly-spaced-bars xylophone, Unevenly-spaced-bars xylophone, Either) were significantly different from baseline probabilities (33.3%), $\chi^2(2) = 14.0$, $p = 0.0009$ (Chi-square test of goodness of fit).

Discussion

If children are using causal reasoning, they should be more uncertain about the structure of the environment when the agent is present, than the inanimate object. Unlike adults, this was not the case for children. 6-year-old children did not choose the "could be either one" option more frequently for the animate agent than for the ball. Instead they systematically picked the xylophone that matched the sounds (e.g. unevenly-spaced bars for unequally-timed notes) often in all conditions. In contrast, 4-year-old children chose "could be either one" at a higher rate throughout all four test trials, and did not systematically pick the matching xylophone.

The third option of "could be either one" may not have been fully comprehensible to the younger children in our sample. The "either" option was presented as a question in a blank box and placed between the two other distinctive physical contexts (evenly-spaced xylophone, unevenly-spaced xylophone). The question mark in the box might have been misinterpreted as an indicator of uncertainty (e.g., I don't know), since the movements behind the curtain were not shown throughout the study. A follow-up version of this study could use a different measure to ask children about the "either one" option, such as putting illustrations of each kind of xylophone into a single picture or asking two separate yes/no questions about each of the xylophones rather than presenting a single answer choice.

Overall, however, the finding that 4-year-old children do not pick the matching xylophone, and instead express uncertainty more often, is consistent with the idea that 4-year-old children are not engaging in causal reasoning in this task. It is thus possible that 4-year-old children cannot or do not engage in causal reasoning about sounds, consistent with the findings from 4-year-olds in Experiment 1.

6-year-old children in contrast may be using causal reasoning to answer that the environment matches the sound, even in the agent condition. Many adults show this pattern as well, and this may be expected even under causal reasoning. The matching sound pattern is the easiest to produce, and people expect other agents to act in ways that minimize effort while maximizing reward (e.g. Jara-Ettinger et al 2015). Since participants do not have any reason to expect that one sequence of tones would be strongly preferred over the other, they may reason that the agent is likely to take the easiest path downhill, rather than making additional effort to change speed and direction to produce the non-matching melody. 6-year-old children may be less meta-aware of their own uncertainty as compared to adults (e.g. Ghetti et al., 2013). This may explain why 6-year-old children did not express uncertainty more often in the agent condition.

General Discussion

Overall, here we find that adults and 6-year-old children (but not younger) engage in causal reasoning about the movements and agents that cause musical sounds. In Experiment 1, participants observed a xylophone being covered by a curtain, heard a sequence of notes, and then were asked to guess whether an agent or inanimate ball had been present behind the curtain.

Adults (Experiment 1a) and 6-year-old children (Experiment 1b) showed causal reasoning, inferring the presence of an animate agent when self-propelled movements were needed to create the musical sounds in that physical context. In contrast, 4-year-olds did not show adult-like reasoning and instead were biased to link orderly sound outcomes (e.g., evenly-timed-note melody) with the presence of an animate agent.

In Experiment 2, we tested whether people's reasoning was flexible, and could be used to infer the physical context in which sounds were produced. Adults again showed clear evidence of causal reasoning. In particular, when an agent was present (vs the ball), and musical notes played, they more often judged that either xylophone could have been behind the curtain – as predicted if they reasoned that the agent could make many possible movements, producing many possible note sequences, while the ball could not. Children did not report greater uncertainty on trials with the agent, potentially due to lesser meta-awareness of their own certainty levels. However, adults and 6-year-old children's judgments were similar in another respect: both most often chose the spacing that "matched" the timing of the sounds. This is in line with causal reasoning, as people expect agents to move in rational, efficient ways (Jara-Ettinger et al., 2016), and these "matching" judgements imply an expectation that the agent, as well as the ball, would take the most efficient path of movement (straight downhill). 4-year-olds, in contrast, performed at chance level. Overall, we believe the most parsimonious interpretation of findings from both experiments together is that adults and 6 year old children (but not younger children) engaged in causal reasoning about the movements and agents needed to cause musical sounds.

Findings with four-year-olds show interesting evidence of developmental change, and suggest that early in life, a different cognitive process links musical sounds with agency. In particular, four-year-olds showed a bias to link orderly outcomes (e.g., evenly-timed notes melody) with the action of an animate agent (Exp 1a), which parallels biases to link order with agency that have been reported for non-musical stimuli (e.g., Newman et al., 2016).

Causal reasoning is likely one process among several that links music with movement and agency. For example, learned associations between musical sounds and people may also play a role (e.g., Launay, 2015). High-level concepts and categories may also guide our reasoning: Once we categorize sounds as "music", we may assume that the sounds were caused by people. However, our data show that by age 6, causal reasoning can overrule these other possible cognitive processes. In our study, neither children nor adults did concluded that all musical sequences were produced by agents (as hypothesized by Launay, 2015). By age 5, children did not systematically expect that more orderly sequences were always associated with agents (as seen e.g., in Newman et al., 2010), and by age 6 children showed causal reasoning similar to adults. Overall, we suggest that causal reasoning is likely an important way by which listeners link music with agency, and that other processes also play a role.

Can we generalize to everyday music listening?

Do people engage in causal reasoning during everyday music listening? In our study, participants were explicitly asked to reason about whether an animate agent or inanimate object was present. Our data therefore serve as a proof of the concept that, when prompted, people can engage in causal reasoning about musical sounds. The extent to which this process occurs implicitly and spontaneously during everyday music listening remains an open question. We hypothesize that causal reasoning occurs during real-world music listening, though perhaps not automatically and not in all cases. It may particularly occur when deeply engaging with music, as in active or creative music listening (Dunn, 1997; Kratus, 2017; Peterson, 2006); or it could be utilized during mental imagery or mental narratives related to musical content (Margulis et al., 2019).

Another way that everyday music differs from our stimuli is in length. Our musical stimuli consisted of sequences of five notes, whereas a typical musical sequence contains a greater number of notes. Nonetheless, the causal reasoning process shown here is easily generalized to longer musical sequences. Indeed, longer musical sequences likely provide even stronger evidence that an agent must be the cause. Producing long, complex musical sequences almost always requires repeatedly changing direction and speed in a self-propelled and intentional way—and this is the case in almost any plausible physical environment. That is, as songs get longer, one would have to posit the existence of increasingly implausible and complex Rube-Goldberg machines in order to think that the notes could be produced without an agent (e.g., by an inanimate object rolling downhill). Thus, particularly for longer musical sequences, causal reasoning should allow the listener to infer that an agent was the most likely cause of the musical sounds, even when they are uncertain of the exact nature of the musical instrument or the physical context of production.

Reasoning about music and beyond

Why is music linked with movement and social concepts? Many influential accounts focus on the role of particular perceptual attributes of sound or on low-level perceptual-motor links, which have been well-established (Cannon & Patel, 2021; Juslin & Laukka, 2003). The current findings suggest that high-level cognition may also play a role. We found evidence that people can use causal reasoning to infer that an animate agent is the cause of musical sounds.

Participants make different inferences about whether an agent is present to cause each sound sequence depending on whether self-propelled movement is required to cause the sounds. When the visual context implies that the notes could be produced in an alternative way, this “explained away” the agent such that people were less likely to infer that an agent was present to cause the sounds.

These findings support the idea that people can reason about the kind of force—animate, or inanimate—causing the musical sounds they hear. This kind of causal inference is a domain-general ability that is not unique to musical sounds, and it likely occurs in a similar way for non-musical environmental sounds (Gerstenberg et al., 2018; Traer et al., 2020). However, the finding that people can reason about agents as causes of musical sounds in particular helps to explain why listening to music activates concepts of movement and animate agency: These abstract ideas may be inferred through causal reasoning. In particular, people may engage in a process of joint inference about the causes of sounds (including music), jointly inferring the type of movement required to cause the sounds and the type of force that could cause those movements (e.g., a person or the wind). To make and test nuanced quantitative predictions about people’s reasoning, future work may model this hypothesized reasoning as joint Bayesian inference (e.g., Baker et al., 2017).

The current work focused on the use of timing cues in causal reasoning, rather than other features of sound (e.g., pitch, timbre). For non-musical environmental sounds, these other features also constrain inferences about possible causes (e.g., Gaver, 1993), and people use multiple acoustic features to make causal inferences about physical-mechanical sound sources (e.g., intensity and reverberation, Traer et al., 2020; see also Agrawal & Schachner, 2022;

Cusimano, Hewitt, Tenenbaum & McDermott, 2019). Future work should investigate the extent to which people integrate and flexibly use multiple cues to make causal inferences about musical sounds and whether causal reasoning about music resembles causal reasoning regarding other environmental sounds.

Causal reasoning about music in situations with greater uncertainty

In Experiment 1, we specified the context in which the sound was produced (the scene with the xylophone) for participants. This allowed us to manipulate the visual context and test its effects on people's judgements. However, people often hear music without seeing the context in which it is played. Do people reason about the role of animate agents when hearing music without seeing the physical context?

We believe that people likely do engage in causal reasoning in this case (though, perhaps not automatically or at all times). Music listening without visual access can be thought of as introducing a greater amount of uncertainty to the causal reasoning process or as starting with a larger hypothetical space. Even with more uncertainty, people can make an inference about the events needed to cause the musical sounds guided by their prior knowledge of how environmental sounds are caused, more generally, and of how musical instrument sounds are created, in particular. People can weigh the likelihood of the musical sequence under a variety of hypothesized causal events and then marginalize over those possibilities (Griffiths & Tenenbaum, 2009). The framework of probabilistic causal reasoning thus applies equally well to cases in which the visual context is not provided. Future work should test the extent to which people engage in causal reasoning spontaneously during everyday music listening.

Developmental Implications

Here we find evidence of causal reasoning only in school-age children, and not younger. Do infants and younger children ever reason about the causes of musical sounds? One possibility is that using implicit tasks (such as looking time) may reveal earlier competencies in causal reasoning in infancy. There are several reasons to expect that younger children and infants may be able to infer that agents are the cause of the musical sounds they hear. From infancy, children differentiate between animate agents and inanimate objects and know that agents have the unique capacity for self-propelled, goal-directed movement (Spelke & Kinzler, 2007; Gelman et al., 1995; Gergely et al., 1995). When infants see an event that requires self-propelled movement, they infer that an animate agent must be the cause (Saxe et al., 2005). In the first months of life, infants link sounds with specific events, for example, expecting to hear a sound when they see an impact between two objects or between an object and a surface (Bahrick, 1983, 1988; Kopp, 2014; Lewkowicz, 1992, 1994; Spelke, 1979). Thus, even infants may engage in causal reasoning about sounds and realize that certain sounds could not be produced without an agent's intervention. Future work should investigate the development of the ability and tendency to infer social causes of musical sounds in childhood.

Alternatively, these tasks may reveal that from early in life, infants link music with agency due to its orderliness (e.g. Newman et al., 2010). Either of these explanations may provide an explanation for infants' early interest in instrumental music: Infants may infer the presence of an agent causing the musical sounds they hear. This would be particularly interesting because in modern times, young children often have more experience with recorded music—for which no agent is present causing the sound—than with live instrumental performance (Mehr, 2014). Thus, any tendency to link music with agents would be in spite of children's early experience with recorded music, which may provide counter-evidence for the causal connection of music to agents.

Conclusion

Overall, the current work shows that adults and 6-year-old children use causal reasoning to infer whether an agent or inanimate object produced musical sounds, integrating information from the visual context and the timing of musical sounds. Such rational inferences about the cause of music were observed when they were asked to infer whether an animate agent or object was present, and thus whether the movements of animate agents (vs. an inanimate objects) were responsible for creating the music. In addition, adults and 6-year-old children were able to infer the physical context (the structure of the xylophone) from musical sounds, and adults expressed more uncertainty when an agent was present – as predicted if they reasoned that the agent could move and stop in a self-propelled way, and thus be able to produce either sound sequence in either physical context. 4-year-old children showed no evidence of causal reasoning about musical sounds, and instead more often inferred that an agent was present during orderly (vs. less-orderly) musical sequences. Overall, these findings provide evidence that by school age, musical sounds can be linked with movement and social concepts through high-level cognition, as a form of rational causal inference about the movements and agents required to cause the sounds.

Acknowledgements

Experiment 1a in Chapter 2 has been previously published in Kim & Schachner (2021) in the Proceedings of 43rd Annual Meeting of the Cognitive Science Society. The dissertation author is the first author of that paper.

Thank you to Heidy Davila, Kaye Han, Joanna Jain, Angela Mak, Jubilee Izzy Shroff and Alyssa Quy for their assistance with data collection. We also thank Madison Pesowski for her input on the study design. Finally, we thank the participants for sharing their time and attention in making this research happen.

CHAPTER 3. PRESCHOOLERS' GESTURE PRODUCTION PREDICTS ANALOGICAL TRANSFER

Learning requires more than acquiring new knowledge; it depends on the application of that knowledge to new contexts. Analogical reasoning—the ability to go beyond perceptual cues to identify abstract structural similarity between events—is therefore of particular significance for learning, since it enables knowledge transfer. For example, if a Windows user switches to a Mac, the ability to map functions from one system to the other will facilitate learning how to use the new computer. For instance, recognizing that the function of the *Command* key on a Mac is analogous to that of the *control (CTRL)* key on a PC will allow a novice Mac user to apply their prior knowledge of their PC to execute specific tasks.

In order to successfully draw this analogy, the learner must ignore salient surface differences (e.g., between the words, symbols, and placement of the keys) and recognize the abstract similarities between their functions (e.g., Brown et al., 1986; Richert et al., 2011). This type of task often poses a challenge for children, who tend to privilege perceptual over structural similarity (e.g., Gentner & Toupin, 1986; Gentner, 1988; Gentner & Rattermann, 1991; Holyoak et al., 1984).

A variety of minimal scaffolds have been used to support children's recognition of abstract structure, including prompts to compare or explain (e.g., Christie & Gentner, 2010; Walker & Lombrozo, 2017) and providing relational language (e.g., Christie & Gentner, 2014). In one classic study by Brown and colleagues (1986), 3- to 5-year-olds were told a story in which a protagonist solves a problem (i.e., transferring objects across a barrier) by repurposing a flat object to create a hollow tube and passing the objects through (Brown et al., 1986). Four- and 5-year-olds successfully applied this abstract solution to a new story that was structurally similar, but superficially distinct. Critically, however, children's success in this task was mediated by their ability to recall the "goal structure" (i.e., the protagonist, the goal, the obstacle, and the solution) of the problem in the original story. The authors concluded that the key factor predicting transfer was whether children represented the underlying abstract structure. Here we consider whether children's *gestures* may be linked to their success in a similar analogical reasoning task.

Several recent proposals suggest that representational gestures (i.e., gestures that reference information related to the content of concurrent speech or thought; McNeil, 1992) facilitate knowledge transfer by selectively schematizing some features over others (e.g., Kita et al., 2017). Indeed, prior work shows that observing co-speech gestures (Guarino & Wakefield, 2020; Guarino et al., 2021) and receiving training to produce gestures (Novack et al., 2014) supports abstraction during early learning. However, there are few studies exploring whether children's *spontaneous* gesture production plays a similar role (e.g., Church & Goldin-meadow, 1986). Below, we review existing theoretical and empirical work that provides support for this hypothesis and describe our novel approach.

Gesture Production Facilitates Abstract Reasoning

According to the *gesture-for-conceptualization hypothesis* (Kita et al., 2017), gestures serve to schematize information for both communication and thinking by stripping away superficial details. This, in turn, shapes the various functions of gesture:

First, gestures *activate* relevant knowledge that supports reasoning and facilitates problem-solving (Church & Goldin-Meadow, 1986). For example, when children are asked to explain a Piagetian conservation task, those who are allowed to use their hands tend to activate spatial features (e.g., height, width), which scaffolds their understanding (Goldin-Meadow & Wagner, 2005).

Second, gestures *manipulate* abstract concepts, often independently from the particular context in which they appear. For example, 4- to 6-year-old children who were taught to use gestures as a strategy for solving mental rotation tasks performed better than those who were trained to physically move objects (i.e., Wakefield et al., 2019). These findings suggest that isolating structural information from the visual and physical details of the problem fosters abstract reasoning.

Third, gestures help learners to *package* relevant information into units that can be flexibly recruited and combined across different contexts. For instance, when children were trained to use abstract gestures as a strategy for solving problems using the equality principle (e.g., a+b+c = ___ + c), they were more likely to transfer this concept to novel contexts than those trained to manipulate sets of magnetic number tiles representing each value or simply simulate the grasping action of physical tiles (Novack et al., 2014). This provides additional evidence that gesture production may support knowledge transfer.

Finally, gestures allow learners to *explore* various potential solutions for a problem before they have identified the correct one. In fact, children often produce relevant information in their gestures before it appears in their speech. These gesture-speech incongruities have been interpreted to indicate that the learner is in a transitional period of conceptual change (Alibali, Kita & Young, 2000; Church & Goldin-meadow, 1986; Pine et al., 2004).

Here, we build on this prior work to examine whether children's gesture production may provide a window into their analogical transfer, staring with spontaneous gesture production. As noted above, the majority of prior work examining the relationship between children's gestures and learning has relied on training or explicitly prompting children to use gestures. Far fewer studies have observed children's spontaneous co-speech gestures during reasoning tasks (e.g., Church & Goldin-Meadow, 1986; Iverson & Goldin-Meadow, 1997), and, to our knowledge, no previous studies have observed children's spontaneous gestures during analogical transfer. Findings with adults provide support for the hypothesis that spontaneous gesture production may be related to analogical reasoning. Specifically, Cooperrider and colleagues (2016) demonstrate that adults often spontaneously recruit gestures to represent abstract relationships while explaining causal events (e.g., stock values rising and falling). Indeed, over 70% of the total number of gestures adults produced referenced structural information. In fact, evidence suggests that spontaneous gesture may be a *more effective* scaffold for abstraction than prompts to use co-speech gestures. In one study, adults who received explicit instructions to produce gestures while providing verbal explanations were *less* likely to engage in analogical transfer than those who gestured spontaneously (Hostetter et al., 2016).

The Current Chapter

The current chapter is composed of two experiments. In Experiment 1, we used a correlational design to examine whether there is a significant relationship between children's *spontaneous* gesture production and analogical transfer success. In Experiment 2, we implemented an experimental design to observe the causal relationship between children's spontaneous gesture production and analogical transfer success. We manipulated children's access to gesture (e.g. free hands vs. sitting on their hands) and their exposure to the experimenter using co-speech gestures (i.e., hearing the experimenter retell the story with or without co-speech gestures).

Prior work suggests that training children to use abstract gestures may facilitate abstract reasoning in mathematics (Novack et al., 2014). However, there is also evidence that direct instruction to use gesture may pose additional cognitive load on participants (Hostetter et al., 2016). Specifically, Hostetter and colleagues (2016) found that when adults were *explicitly* instructed to use their hands while explaining how they solved a problem, they performed worse in analogical transfer compared to those that were not instructed to use their hands. To avoid this potential issue, we implement a manipulation to promote gesture production in Experiment 2 by having children watch an adult gesture (Watch-Gesture condition). This decision was based on prior studies by Cooperrider and colleagues (2017), who found that adults who observed abstract gestures that highlighted relevant knowledge in an instructional video later incorporated more abstract structural information in their explanation compared with those who observed iconic gestures. We then compared children's analogical transfer in this group to another group of children who were prevented from gesturing during retelling (Sit-on-hands condition).

Experiment 1

Experiment 1 is an initial study that explores the relationship between children's analogical transfer and their spontaneous gesture production. We consider two questions: First, do children who spontaneously produce gestures perform better on analogical transfer tasks, compared to those who do not? Given that successful analogical transfer hinges on children's recognition of common abstract structure (e.g., Brown et al., 1986), and that spontaneous gestures have been shown to promote abstraction (e.g., Church & Goldin-Meadow, 1986; Cooperrider et al., 2016; Kita et al., 2017; Pine et al., 2004), we predict greater rates of analogical transfer in children who spontaneously gesture when retelling exemplar stories.

Second, does the *type* of spontaneous gestures produced (i.e., solution-relevant vs. solution-irrelevant) correspond to children's success on analogical transfer? Given that spontaneous gestures tend to reflect abstract relationships over superficial details (e.g., Cooperrider et al., 2016), we predict a higher rate of success among children who produce solution-relevant gestures.

Following Brown and colleagues (1986), we presented children with three superficially distinct stories in which the protagonists each shared a common goal and discovered a common solution. The first two exemplar stories were each followed by a retelling task. We included *two* exemplar stories to promote comparison and facilitate abstraction (e.g., Gentner & Hoyos, 2017). The third story presented an analogous problem, and children were asked to provide the solution themselves. We included children aged 5- to 7-, since they are capable of succeeding in similar narrative-based analogical transfer tasks with minimal scaffolding (e.g., Richert et al., 2011; Holyoak et al., 1984).

Method

Participants. Eighty-five 5- to 7-year-olds participated in the study (M = 6.51 years, SD = 0.83, range = 5.0-7.92 years), with half of these children tested in person (n = 43, M = 6.56

years, SD = 0.89, range: 5.0-7.92 years) and the other half tested remotely (n = 42, M = 6.45 years, SD = 0.78, range: 5.17-7.67 years). An additional 9 participants were excluded (2 in person and 7 remote) due to stimuli malfunction (1), experimenter error (2), technological issues (5) or interruption by a sibling (1). Children who participated in-person were recruited and tested at a local museum, an aquarium, or at local preschools, and received small gifts. Children who participated remotely were recruited through the lab database, and received a $5 Amazon gift card.

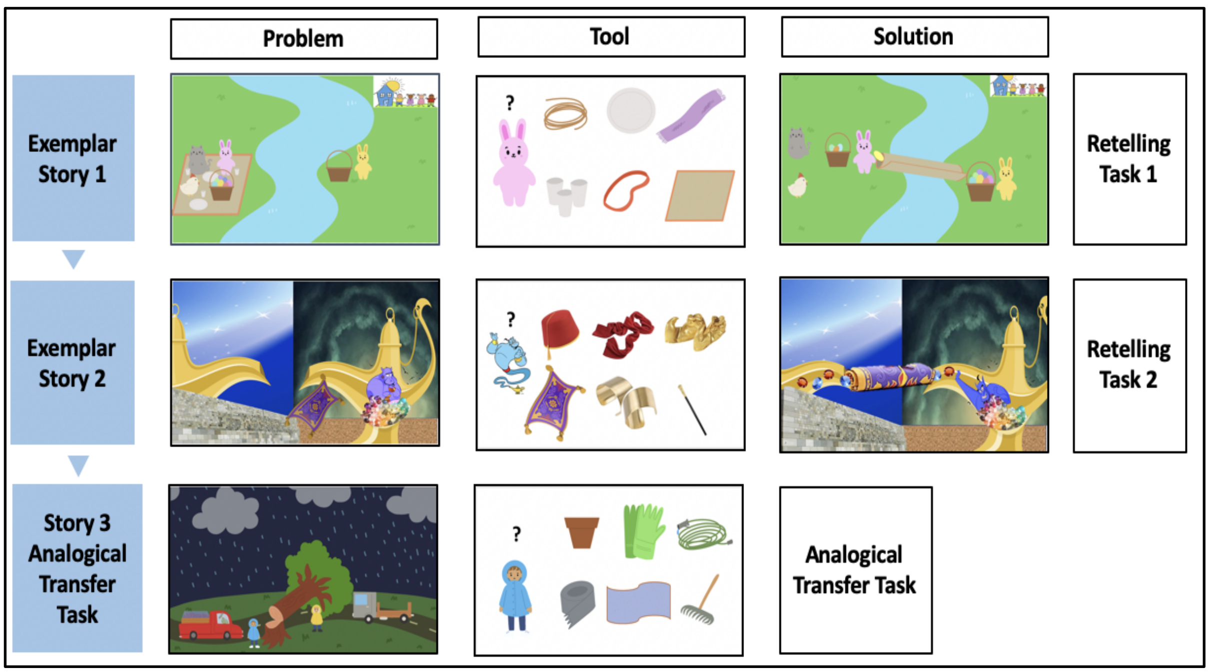

Figure 3.1: Experiment 1 Stimuli. Three stories with a common abstract problem and solution were used. In Exemplar Story 1, a bunny wanted to deliver eggs to the other side of the river. The bunny rolled a picnic mat into a hollow tube, and passed the eggs through it. In Exemplar story 2, a genie wanted to move his jewels to a new lamp across a wall. The genie rolled his magic carpet into a hollow tube, and passed the jewels through it. Exemplar Stories 1 and 2 were each followed by a retelling task. In the Analogical Transfer Task story, a farmer wanted to deliver the cherries across a fallen tree in the road. Children were presented with tools (including the target tool, a flat truck cover), and asked to solve the problem themselves.

Stimuli. Three stories, including two exemplar stories (presented in counterbalanced order), and an analogical transfer task, were presented as PowerPoint slides on a 13-inch laptop screen (in-person) or via an online presentation platform (slides.com) during a Zoom session. Each story included static images of a protagonist, a set of tools, and a barrier. Stories were modified versions of those used in Brown et al. (1986) (see Figure 3.1).

The first few slides of the exemplar stories presented the protagonist's goal (i.e., a bunny wants to deliver painted eggs to children on the other side of a river; a genie wants to move his jewels to a new lamp across a wall), followed by a description of the barrier (i.e., a river; a wall). The next slide displayed six tools that the protagonist could use to solve the problem, including the target object (i.e., a picnic mat; a magic carpet). The exemplar stories also included solution slides demonstrating how the protagonist solved the problem. On the first slide, the protagonist rolled a flat material (the target object) into a hollow tube. Then, on the second slide, the protagonist passed the objects (i.e. the painted eggs; the jewels) through the tube to the other side of the barrier.

The third story was used for the analogical transfer task. The first few story slides presented a new character with an analogous goal (a farmer wants to transfer the cherries in his truck across the road, which is blocked by a fallen tree). The next slide displayed six tools, including the target object (a flat truck cover). Since children were prompted to generate the solution themselves, no solution slides were provided.

Other materials included a photograph of a child, which was used for the retelling tasks after the exemplar stories.

Procedure. *In-person version***.** Children were tested, one-on-one, in a quiet area of a museum or preschool classroom. All materials were presented on a laptop screen. Participants were told that they would be listening to three stories, and that they would be asked to answer some questions about the stories.

The experimenter read the first exemplar story, which introduced a protagonist and his goal, followed by a description of the barrier (see Figure 3.1). After describing the problem, the experimenter proceeded to the *tools slide*, and labeled each of the six tools available, while pointing to the objects on the screen. The experimenter briefly prompted the participant to consider what the protagonist could do to solve his problem, but did not give them sufficient time to answer. Instead, the experimenter proceeded to the solution slides and completed the story.

In the *retelling task* for the first exemplar story, the participant was asked to look at a picture of a gender-matched child attached to the video camera, and were told that the child would be watching this video later on. With the laptop screen closed, the participant was prompted to tell the child what happened in the story and how the protagonist solved the problem. The experimenter said: “This is a picture of another child named Alex. Alex won’t get to hear this story. Instead, he/she will be watching your video! Can you tell Alex what happened in the story and how [character] solved the problem?” Participants were not prompted to gesture, but the study session videos were coded offline to record children’s speech and spontaneous gestures.

Afterwards, the experimenter began the second exemplar story, which presented an analogous problem and solution (see Figure 3.1). Again, children were prompted to look at the camera and retell the story events for another child to watch.

The third story was used to assess analogical transfer. After presenting the analogous problem, the experimenter told the child that they will be solving the problem themselves this time. The experimenter labeled the available tools that the protagonist could use. Then, the experimenter asked the child how the protagonist could solve his problem using one of the tools (e.g., "Can you tell me, how can Farmer Jones move his cherries across the big fallen tree?"). Here, we coded whether children used a similar solution from the previous stories.

Modification for the online version. The online version of the experiment was conducted on Zoom. Stories were presented using Slides.com. Before beginning the session, parents were instructed to set up their webcam at an angle that could capture both the child's face and hands. Although the procedure for the online version was identical to the in-person version, the narration for each story was pre-recorded. In place of pointing, an animated arrow indicated each object as it was introduced.

Coding. Children's spontaneous gestures were coded from the video by a naive researcher. Children who produced spontaneous gestures while retelling either of the exemplar stories were coded as 1, and those who did not gesture at all were coded as 0. This variable was used to form post-hoc groups (*gesture*, *no-gesture*) for analysis, which served as a binary predictor variable in logistic regressions.

Among the participants who produced gestures, we also coded whether those gestures were relevant to the analogical solution. Gestures that were spontaneously produced while describing the solution (e.g., rolling, passing through) were given a score of 1 (solution-relevant), and all other gestures were given a score of 0 (solution-irrelevant). Solution-relevant gestures included, for example, rotating both hands or placing both fists next to each other and tilting the wrists forwards and back (for rolling), and then moving one hand to the other (for passing objects through). Solution-irrelevant gestures included non-referential gestures (e.g., finger tapping for particular words), or gesturing superficial content that was not central to the abstract solution (e.g., using both hands to simulate lowering the picnic mat down to the ground).

Children who produced solution-relevant gestures at least once during either of the retelling tasks were categorized into the solution-relevant gesture group. Children who only produced solution-irrelevant gestures were categorized into the solution-irrelevant gesture group.

Children's performance on the analogical transfer task was coded based on whether they described the abstract solution (i.e., rolling a flat material into a tube and passing the objects through). A minority of children who only mentioned the first part of the solution (i.e., rolling the flat material into a tube) in their descriptions were also coded as correct. The participant videos were coded independently by two researchers, and the overall inter-coder reliability on children's gesture production and analogical transfer ratings was 84.1%. Any discrepancies were resolved in discussion with a third researcher.

Results

Chi-square tests indicated that the testing medium (in person vs. online) was not related to children's spontaneous gesture production, $\chi^2(1) = 0$, $p = 1$), or children's analogical transfer performance, $\chi^2(1) = 0.10$, $p = 0.75$. We therefore combined the two participant samples for all analyses.

As noted above, children were categorized into two groups: those who produced spontaneous gestures during at least one of the two retelling tasks (*gesture* group, $n = 48$ out of 85, 56.5%) and those who did not (*no-gesture* group, $n = 37$ out of 85, 43.5%). The overall analogical transfer rate in our sample (23%) was low, but consistent with findings from prior work in this age group when no additional hints or scaffolding are provided (e.g., Brown et al., 1986; Holyoak et al., 1984; Kim & Choi, 2003). Among children in the *gesture* group, 16 of 48 succeeded at the analogical transfer task (33.3%), whereas in the *no-gesture* group, only 4 of 37 participants succeeded (10.8%). A chi-square test of independence indicated a significant relationship between children's spontaneous gesture production and their tendency to engage in analogical transfer, $\chi^2(1) = 4.71$, $p = 0.03$. This finding was confirmed using logistic regression, Wald-Z-test; $\chi^2(1) = 5.4$, $p = 0.02$. These results support our main hypothesis that children who produced spontaneous gestures would perform better on the subsequent analogical transfer task, compared to children who did not produce gestures.

Age did not significantly predict gesture production, Wald Z-test; $\chi^2(1) = 0.27$, $p = 0.60$ or analogical transfer $\chi^2(1) = 0.01$, $p = 0.92$. Adding age as an additional variable to our main

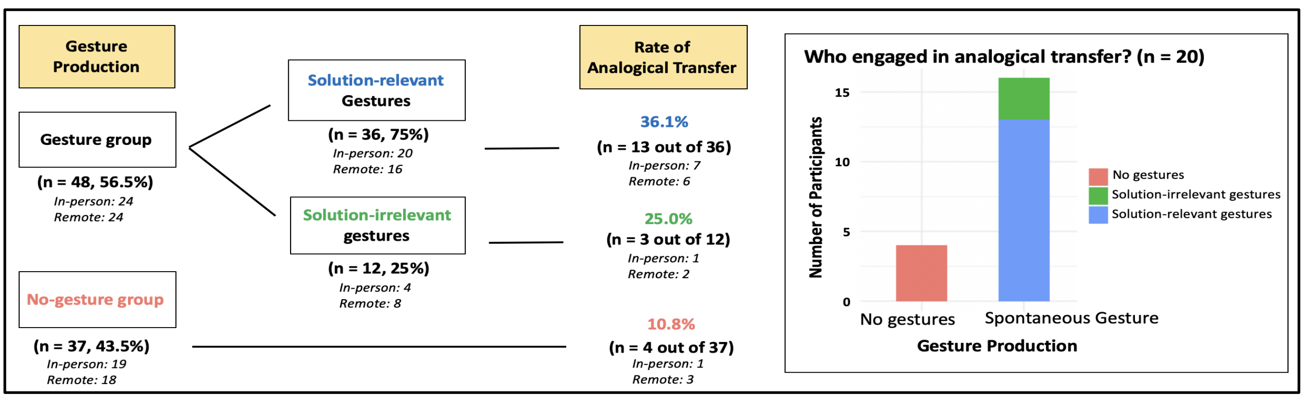

Figure 3.2: Experiment 1 Results. The diagram on the left shows the proportion of participants who gestured during the retelling tasks, the proportion of participants who produced solution-relevant and solution-irrelevant gestures, and the rates of transfer for each group. The chart on the right shows the number of participants in each group who engaged in analogical transfer.

model did not perform better than the model alone; $\chi^2(1) = 0.32$, $p = 0.57$ (nested model comparison).

Relevance of gesture to the goal-structure. Thirty-six of the 48 participants in the gesture group produced solution-relevant gestures (75.0%), and 13 of these 36 participants also engaged in analogical transfer (36.1%). In contrast, 12 out of 48 participants produced only solution-irrelevant gestures (25.0%), and only 3 out of these 12 participants engaged in analogical transfer (3/12, 25.0%, see Figure 3.2). Despite this numerical difference in the rate of analogical transfer between groups, this was not statistically significant, $\chi^2(1) = 0.13$, $p = 0.72$.

Individual differences in speech. As a proxy for individual differences in engagement, we analyzed the total number of words produced during both retelling tasks, excluding any verbal disfluencies ($M = 49.08$). Although we found a significant correlation between word count and children's spontaneous gesture ($M_{gesture} = 58.30$, $M_{no\ gesture} = 37.10$; $r = 0.31$, $p = 0.004$), word count was not correlated with participants' success on analogical transfer ($r = 0.14$, $p = 0.20$) or their tendency to produce relevant gestures during retelling ($r = 0.08$, $p = 0.43$). Finally, adding word count as an additional predictor to our main model did not increase the explanatory power of the model; $\chi^2(1) = 0.31$, $p = 0.57$ (nested model comparison).

Discussion

In Experiment 1, we aimed to establish a relationship between children's spontaneous gestures and their subsequent analogical transfer. Our results provide preliminary evidence for this: Children who produced gestures when retelling one or both of the exemplar stories were significantly more likely to abstract a common solution to a superficially distinct target story. We also sought to determine whether the type of gestures produced mattered for transfer. Although we observed a numerical difference in the rates of analogical transfer, with a higher rate of success in children who produced solution-relevant gestures, this difference was not statistically significant. Since this null result may have been due to the small sample size of the *gesture* group, future work will replicate this study with a larger sample.

Critically, although we find evidence for a relationship between children's spontaneous gesture production and analogical transfer, the correlational design of Experiment 1 cannot establish causality. Indeed, several other variables may have plausibly contributed to this relationship. Specifically, children who were more attentive or engaged during storytelling may have been more likely to produce gestures during retelling and also more likely to succeed on the analogical transfer task. As a proxy for engagement, we analyzed the total number of words produced by each child during the retelling tasks. Although word count was correlated with gesture production, it was *unrelated* to success on analogical transfer. It is therefore unlikely that increased transfer in the gesture group can be explained by differences in verbal production.

Experiment 1 provides initial evidence for the relationship between children's spontaneous gestures and their analogical transfer, suggesting that gesture may provide a window into early abstract reasoning. Although this was one of the first studies to establish this relationship, Experiment 1 was limited by its nature being a correlational design and not being able to establish causal directions between the two variables. Therefore, in Experiment 2, we examine the causal relationship between children's gesture production and analogical transfer by using an experimental design. We introduce two new conditions, one that promotes the usage of gestures (e.g., Watch-Gesture condition) and another that constrains children from using their hands (e.g. Sit-on-Hands condition). These interventions will not only help to establish whether gesture *supports* relational transfer in our task, but will also shed light on the potential differences between spontaneous and prompted gesture in children (see Hostetter et al., 2016 for related work in adults).

Experiment 2

In Experiment 2, we dive deeper into the cognitive mechanism underlying the relationship between gesture production and analogical transfer to examine the causal direction. Will promoting gesture increase children's analogical transfer? Will analogical transfer be hindered if children are prevented from gesturing during retelling? We designed Experiment 2 to compare two conditions that manipulate children's access to gesture: one in which children observe the experimenter produce co-speech gestures and are encouraged to gesture themselves (Watch-Gesture condition), and one where children are asked to sit on their hands during their retelling tasks, preventing access to gesture (Sit-on-Hands condition).

We test three predictions: (1) Children who are encouraged to use gesture will be more likely to engage in analogical transfer than children who are prevented from using gestures; (2) Children who produce gestures will be more likely to engage in analogical transfer than those who do not; (3) Children who produce gestures that are relevant to the solution will show a higher rate of analogical transfer than children who only include gestures irrelevant to the solution and children who do not gesture at all.

Method

Participants. 121 five- to seven- year-olds (M_{age} = 77.8 months, SD = 11.12, range = 60.1 to 95.7 months) participated in Experiment 2. This sample size was determined by a power analysis on the main effect of our independent measure on the dependent measure from Experiment 1. As in Experiment 1, participants were recruited in-person (n = 84) and remotely (n = 37) evenly across the two different conditions (Sit-on-Hands, Watch-Gestures). Participants who were tested in person were recruited from local museums, preschools, or were invited to the lab. For participants who were tested remotely, the Zoom video call platform was used. Fourteen additional participants were tested but later excluded due to technical difficulties (2), failure to complete some trials (5), parent interference (1), producing gestures in the Sit-on-hands condition following reminders not to (3), or difficulties coding children's speech due to masking (3).

Design. In a between-subjects design, participants were randomly assigned to either the Watch-Gesture condition (n = 61) or the Sit-on-Hands condition (n = 60). The age range of participants were matched across the two conditions (Watch-Gestures condition, Mean age = 77.7 months; Sit-on-Hands condition, Mean Age = 78.0 months) and equal across the two testing methods (e.g., In-person, Mean age = 77.3 months; remotely on Zoom, Mean = 79.0 months).

Stimuli. Videos and images used in Experiment 2 were identical to that of Experiment 1. In addition, memory questions were added at the end of the analogical transfer task as a measure of individual differences in engagement. For each of the two exemplar stories, seven memory questions were added (14 in total) to ask content that were relevant to the goal structure (e.g., the problem of the protagonist, the obstacle, the tool for solving the problem, and how it was used) or irrelevant to the goal structure (e.g., (hair) color of the protagonist, other characters in the story, and the other tools that were available to the protagonist).

Procedure. The procedure in Experiment 2 was identical to that of Experiment 1 up until the end of the solution video for the first exemplar story. Before guiding the participant to the retelling task, the experimenter told the participant that they will be demonstrating how to retell a story before the participant's turn (e.g., "Now, I'm going to tell you what happened in the (protagonist)'s story, and then it will be your turn to retell the story.").

During the experimenter's retelling phase, the experimenter followed a script that included the four goal-structure elements from Brown et al., 1987 (protagonist in a context, goal, obstacle, solution). In the Watch-Gestures condition, the experimenter used abstract gestures that also represent the elements relevant to the structure of solution to the problem that the protagonist had, such as the two a group of small goal objects to be transferred (e.g., cherries, eggs, jewels), the obstacle (e.g., the river, the wall, the fallen tree), the two sides of the obstacles (e.g., two sides of the river; the farm and the market, the two lamps) and how the protagonist solved the problem by rolling a flat material (e.g., a picnic mat, a truck cover, a carpet) into a tube and passing the goal objects to the destination (See Figure 3.3 below).

After the experimenter retold the first exemplar story, the child was instructed to retell the story to the picture of a child (Alex), as in Experiment 1. At the end of the prompt for the retelling task, different prompts were added depending on the condition. In the Watch-Gesture condition, the experimenter encouraged the participant to use hands just like they did (e.g., "(…) And when you do so [retell], just like what I did, you can use your hands!"). In the Sit-on-Hands condition, the experimenter instructed the participants to sit on their hands (e.g., "(…) And when

	Example: The Bunny Story			
	Context	**Goal**	**Obstacle**	**Solution**
Experimenter Retelling Script	The little bunny was coloring **some eggs** at a picnic with his friends by the river.	He has to deliver his colorful eggs from **one side** of the river to **the other side** of the river for the children	The river is **very large** and there aren't any bridges or boats for them to cross.	So, the little bunny **rolled** the picnic mat into a tube and **passed the eggs** through it to his friend on the other side of the river.
	Context	**Goal**	**Obstacle**	**Solution**
Watch-Gestures condition (Co-speech gestures)	**Some eggs**: Both hands together in a C-shape like a picnic basket	**One side:** Left hand open, flat (knife shape), palm facing the right **The other side:** Right hand open flat (knife shape), palm facing the left	**Very large:** Both hands open flat (knife shapes), palms facing each other to show the width of the river	**Rolled:** both hands, palms down in C-shapes, moving up and down at the rest **passed the eggs:** left hand open, palm facing the right, snap at the wrist to move toward the right
	Context	**Goal**	**Obstacle**	**Solution**
Sit-on-Hands Condition	Speech only	Speech only	Speech only	Speech only

Figure 3.3: An example of the prompt and gestures for the experimenter's demonstration of retelling (the Little Bunny story). In the Watch-Gesture condition, the yellow-highlighted words were accompanied by the gestures in the picture.

you do so, we will be sitting on our hands!"). Then participants continued to listen to the second exemplar story. Then, the second retelling phase was identical to that of the first retelling phase and was composed of the experimenter's demonstration of retelling followed by the participant's retelling task. In the Watch-Gestures condition, similar but perceptually different gestures were used for the key elements of the story to prevent children from merely copying the experimenter's gestures. The rest of the study procedure was identical to that of Experiment 1, followed by the third story video, an analogical transfer task. Then, after the transfer task the memory questions were asked. The experimenter told the child that they will be asking some questions about the first two stories that they have heard and asked the child to try their best to answer them. The order of the memory questions followed the order of the two exemplar stories. Within each story, the experimenter followed an identical order of questions that roughly alternated between contents that were relevant to the goals structure (e.g., “What problem did the [protagonist] have?”; “What was in his way?”; “What did he use to move the [small objects] across?”, “How did he use it?”) and contents that were not relevant to the goal structure (e.g., “What was the (hair) color of the [protagonist]?”; “Who else was in the story?”; “What were the other tools that the [protagonist] had?”). When the child did not have an answer to the question the experimenter repeated the question. If the child still did not have an answer or said “I don’t know” then the experimenter moved on to the next question.

Coding of memory questions. Memory questions for the two exemplar stories were used as a measure of potential individual differences in engagement. Each memory question was coded from incorrect (=0) to correct (=1) with a half-score (=0.5) for partial answers. The accuracy scores were summed by the type of the questions, relevant to the goal structure (4 questions for each of the two stories, 8 as the full score) or irrelevant to the goal structure (3 questions for each of the two stories, 6 as the full score). The proportion of accurate responses were calculated out of the full score for each type of the questions and also by the total score (e.g., out of 14 as the full score). Six participants that had one or more questions that was hard to code due to the background noise were removed from this analysis.

Results

A preliminary analysis was conducted to confirm that there were no effects of testing methods and/or age on the rate of analogical transfer above and beyond condition and gesture production. A multiple regression model with transfer success as the outcome variable and gesture production, condition, child’s age in months, and testing method (in person vs. on zoom) as the predictor variables (Model 1) was used. We found no effect of testing method (β = -0.37, p = 0.46) or child's age in this relationship (β = 0.03, p = 0.14). The same pattern was found when a child's age in months was replaced with an ordinal variable age bin (5-,6-, and 7-year-olds, β = 0.25, p = 0.30). Additional chi-square tests of independence confirmed no relationship between the testing method and the children's analogical transfer performance (χ^2 (1) = 0.00003, p = 0.98), between the testing method and gesture production ($\chi^2(1)$ = 0.64, p = 0.42), between the child's age and gesture production ($\chi^2(2)$ = 0.47, p = 0.79), or between the child's age and analogical transfer success ($\chi^2(2)$ = 2.10, p =0.34). Therefore, the two testing methods and the child's age groups were combined for further analysis.

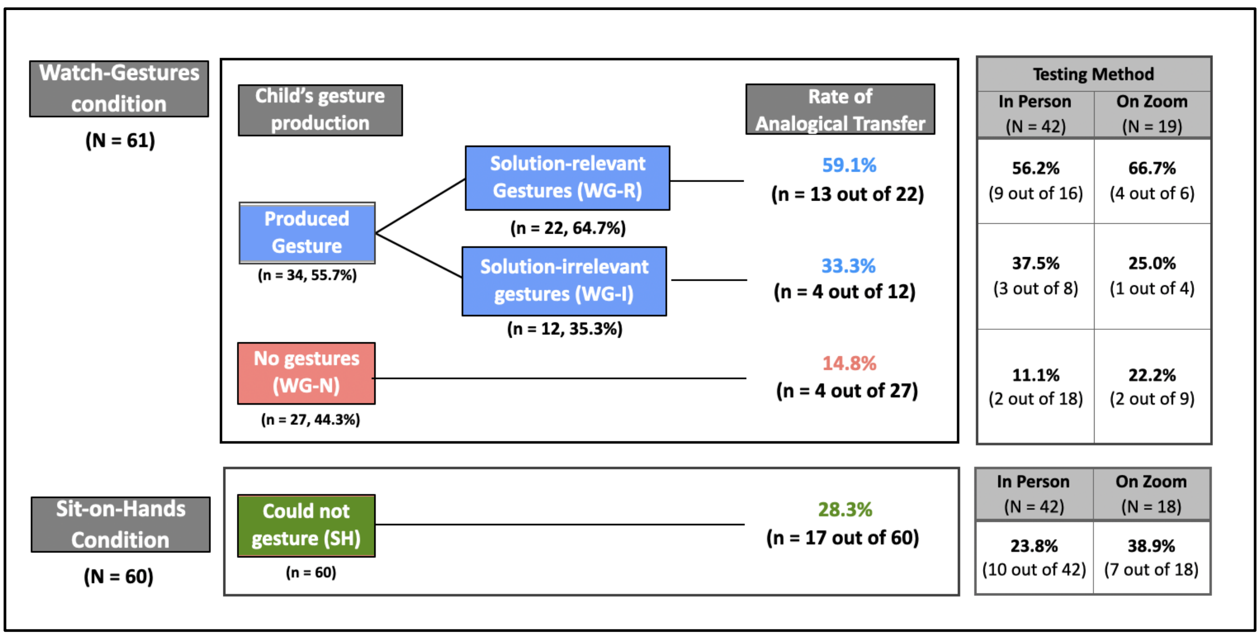

Figure 3.4: Results of the Watch-Gesture condition and Sit-on-Hands condition from Experiment 2. The table on the right shows the breakdown of the number of participants by the testing methods (in-person vs. remotely on Zoom).

Promoting and constraining access to gesture. The Sit-on-Hands condition and the Watch-Gesture condition were compared to one another to examine whether access to gesture predicted success in analogical transfer. We did not find a main effect of condition (*WG* vs. *SH*) on children's analogical transfer success rate (WG, 34.4% vs. SH, 28.3%; $F(1,119) < 1$, $p = 0.95$, one-way ANOVA; $\chi^2(1) = 0.28$, $p = 0.59$; $\chi^2(1) = 0.28$, $p = 0.60$, Chi-square test of independence). In the Watch-Gestures condition (*WG*, n = 61), three sub-groups were created: children who did *not* gesture at all during the retelling task (WG-no-gestures, *WG-N, n = 27 out of 34, 64.7%*), children who gestured content *relevant* to the solution (WG-Gestured-relevant-content, *WG-R, n=22 out of 34, 35.3%*), and children who gestured but only surface features that were *irrelevant* to the goal structure (WG-gestured-irrelevant-content, *WG-I, n =12*). The Sit-on-Hands condition (SH, n = 60) was analyzed as one full sub-group (see Figure 3.4). When only the children who produced gestures in the Watch-Gesture condition (n = 34) were compared to those in the Sit-on-Hands condition (n = 60), children who could gesture performed marginally better in analogical transfer than children who did not have access to their hands (50.0% vs. 28.3%, $\chi^2(1) = 3.52$, $p = 0.06$, chi-square test).

Gesture production vs. No gestures. In the Watch-Gesture condition, children who produced gestures (both solution-relevant or solution-irrelevant) showed a significantly higher success rate for analogical transfer compared to those who did not produce gestures (rate of analogical transfer for children who gestured, 50.0% vs. who did not gesture, 14.8%; $F(1,59) = 6.06$, $p = .017$, one-way ANOVA).

Relevance of gesture to the solution. As in Experiment 1, we also looked at whether the content of gesture in relation to the analogous solution structure affected the rate of analogical transfer. Among the 34 children who gestured in the WG condition, the transfer success rate in those who gestured relevant content (WG-R, n = 22) and those who only gestured solution-irrelevant content (WG-I, n = 12) did not significantly differ from each other (59.1% vs. 33.3%; $F(1,32) = 1.88$, $p = 0.18$). There was a significant difference in the rate of analogical transfer across success rate among the three post-hoc groups within the WG condition ($F(2,58) = 4.21$, $p = 0.02$, one-way ANOVA). A post-hoc Tukey test showed that the transfer rate in *WG-R* and *WG-N* group differed significantly (adjusted *p-value* = 0.015). The rate of transfer in *WG-I* group was not significantly different from the other two groups, lying somewhere in the middle.

Watch-Gesture condition vs. Spontaneous Gesture. The relevance of gesture content in the Watch-Gesture condition (WG) was also compared to that of Experiment 1 (NP). A 2x3 factorial design was used for condition (*WG, NP*) and gesture use (no gestures, relevant gestures, irrelevant gestures). There was no significant effect of the 2x3 interaction of condition and gesture use ($p = 0.39$), however there was a significant main effect of gesture use on analogical transfer ($F(2,140) = 6.08$, $p = 0.003$) and a marginally significant main effect of condition on analogical transfer ($F(1,140) = 3.74$, $p = 0.055$).

Individual differences in engagement. The accuracy for the memory questions were used as an additional measure of engagement. We found no significant differences of engagement by the four post-hoc subgroups (SH, WG-I, WG-R, WG-N) when engagement was measured as the accuracy score for all memory questions ($F(3,102) = 1.87$, $p = 0.14$, one-way ANOVA), for only the solution-relevant questions ($F(3,102) = 2.386$, $p = 0.07$, one-way ANOVA), or for only the solution-irrelevant questions ($F(3,102) = 0.37$, $p = 0.77$, one-way ANOVA). There were no difference in child's engagement between the Watch-Gesture condition and the Sit-on-Hands condition (all questions, $F(1,104) = 0.629$, $p = 0.43$; *solution-relevant questions*, $F(1,104) = 0.91$, $p = 0.34$; solution-irrelevant questions, $F(1,104) = 0.067$, $p = 0.80$). Across the WG and SH condition, here were no differences in children's engagement by whether the child produced gestures or not (*all questions*, $F(1,104) = 0.14$, $p = 0.71$; *solution-relevant questions,* $F(1,104) = 0.048$, $p = 0.83$*; solution-irrelevant questions,* $F(1,104) = 0.184$*,* $p = 0.67$). Among children who produced gestures, there were also no differences in children's engagement by whether the child produced gestures relevant to the solution or not (*all questions*, $F(1,56) = 0.083$, $p = 0.79$); *solution-relevant questions,* $F(1,56) = 0.25$, $p = 0.62$); *solution-irrelevant questions,* $F(1,56) = 0.02$, $p = 0.89$*)*.

Discussion

In Experiment 2, we aimed to examine the mechanism underlying the relationship between gesture and success in analogical transfer. When children were exposed to the experimenter's co-speech gestures prior to the retelling task (*WG*), children who produced *any* gestures in the retelling tasks performed better in analogical transfer than children who did not gesture (*WG-N*). Specifically, children who previously gestured *relevant* content (*WG-R*) showed a higher success rate in analogical transfer than children who did not gesture at all (*WG-N*), paralleling the pattern found in Experiment 1. Interestingly, children who were asked to sit on their hands during the retelling task without seeing experimenter's gestures (*SH*) performed better at analogical transfer than those who did not gesture at all even after seeing the experimenter's gestures (*WG-N*). When analyzed together with the full sample in Experiment 1 (*NP*), a main effect of gesture production was observed on the rate of analogical transfer and a stronger effect between groups with relevant gestures and no gestures produced.

Experiment 2 provides evidence that children's gesture production matters more than exposure to an adult's gestures. Children who recruited gestured during the retelling task performed better in analogical transfer in both the Watch-Gesture condition in Experiment 2 as well as the children who were not prompted to gesture in Experiment 1. Our findings add to the current literature that using hands, either trained or not, facilitates children's abstract reasoning in a new context (Goldin-Meadow & Wagner, 2005; Novack et al.,2014). Specifically, what's represented at children's fingertips may bring valuable insights to children's successful analogical reasoning as observed from the significant differences in children who used solution-relevant gestures and those who did not use gestures at all (e.g., Church & Goldin-Meadow, 1986). The manipulation on promoting children to use gestures numerically increased the rate of analogical transfer compared to Experiment 1.

However, there were some limitations in Experiment 2 that interfere with a straightforward interpretation of the results. Specifically, our manipulation intended to promote gesture failed to increase children's production of gesture during retelling. Indeed, the rate of gesture production in the Watch-Gesture condition closely mirrored those produced following no prompts in Experiment 1 (55.7% of children gestured in WG; 56.5% children gestured in Experiment 1). Although this pattern suggests that the child's gesture production is a stronger predictor of analogical transfer than observing adult's gesture, it also suggests that our manipulation may have been too weak to detect an effect.

Given that not all children in the Watch-Gesture condition produced gestures, it is difficult to draw meaningful comparisons between conditions. By only comparing children who produced gestures to those in the Sit-on-Hands condition, we may be selecting for individuals who have greater cognitive resources that can effortlessly recruit co-speech gestures as they produce speech. Interpretation for children in the Sit-on-Hands condition is limited since we would not know how much the child would have gestured if they had access to their hands. Still, our measure on the memory questions suggests that no one post-hoc group was more engaged in the experiment session than one another. We found no difference in individual engagement across the different conditions, child's gesture production, or whether the gestures were relevant to the goal structure.

General Discussion

In Chapter 3, two Experiments establish a significant relationship between children's gesture production and analogical transfer and examine the underlying cognitive mechanisms. In Experiment 1, we observed that children's spontaneous gesture production (No-Prompt condition) was significantly and positively correlated with their performance in analogical transfer task. In Experiment 2, we compared two conditions where children's gesture usage was either promoted (Watch-Gesture condition) or constrained (Sit-on-Hands condition). In terms of gesture production, about the same proportion of children gestured (vs. did not gesture) in both Watch-Gestures condition and in Experiment 1 where children did not have prompts to produce gestures. In terms of analogical transfer, a main effect of gesture usage was confirmed in both Experiment 1 and 2 (for children who had access to gestures) with a greater effect found for gestures that are relevant to the analogical solution shared among the exemplar stories.

The spontaneity of gesture production

We can conclude that producing gesture has a greater effect on analogical transfer than observing an adult's gestures. Our subtle manipulation of promoting gestures in this condition was based on the previous studies that adults may be cognitively overloaded if they were *explicitly* required to use their hands during an explanation task (Hostetter et al., 2016), but merely *observing* other's co-speech gestures facilitates abstraction (e.g., Guarino & Wakefield, 2020; Guarino et al., 2021). Nevertheless, this manipulation was not strong enough to increase the overall rate of children's gesture production in the Watch-gesture condition. Half of the children in this condition failed to gesture which may introduce possible effects of individual differences. Still, we found that children's engagement did not differ by condition or by child's gesture production, potentially moving away from the assumption that individual differences on engagement would explain children's gesture production.

Indeed, although all children in the Watch-Gesture condition saw the experimenter retell the story with co-speech gestures, it was up to the child on whether they would actually recruit their hands during their retelling task. Therefore, if there are any cognitive effect of gesture production on the analogical transfer success, the child's spontaneous recruitment of hands for processing structural information might play a central role (e.g., even at the absence of any prompts for gesture, Church & Goldin-Meadow, 1986; Goldin-Meadow & Wagner, 2005) than being verbally prompted to use gestures (see Hostetter et al., 2016 for results with adults).

Activating and Manipulating structural information

Experiment 2 provides evidence that being able to recruit hands may play an important role in revisiting the abstract structure of the analogous contexts in children. First, we confirmed that children who produce gestures in the retelling task were more likely to succeed in the later analogical transfer task than children who did not produce gestures in both Experiment 1 and 2, with or without prompts for using gestures. Our results provide evidence to the gesture-for-conceptualization hypothesis (Kita et al., 2017) which proposes four functions of gestures including that gestures *activate* knowledge relevant to problem-solving and *manipulate* abstract concepts by schematizing structural information away from the content-specific details. In our study, our retelling prompts were specifically designed for children to *activate* the abstract concepts (e.g., "Can you tell Alex what happened in the story and how the protagonist *solved the problem?").* Therefore, gestures produced at the retelling stage may provide a window to when relevant knowledge is activated at children's fingertips online as they verbally walk through what's on their mind (e.g., see an example of the conservation task in Church & Goldin-Meadow, 1986).

Moreover, our analogical transfer task provided a novel context where children could apply the analogous solution, ideally by *manipulating* the abstract concepts from the two exemplar stories. Based on our preliminary observations, gesture production were observed at the analogical transfer ask even for children who did not use gestures during the retelling task (in Watch-Gestures condition) or children who were previously asked to constrain from using their hands at the retelling tasks (Sit-on-Hands condition). Although it is hard to tease apart whether gestures activated *or* manipulated the abstract concepts, our observations of gesture production in both the retelling tasks and the analogical transfer tasks provides evidence to the cognitive role of gesture production in abstract reasoning. For a systematical observation of whether children readily recruit their hands to revisit structural information, future studies may add an another task to the Sit-on-Hands condition to observe children's gesture usage when they have access to their hands.

Gesture production as a window to early analogical reasoning

These experiments may have important implications for research on the development of analogical reasoning. As noted in the introduction, children's gestures have been shown to reveal latent concepts that are not yet manifest in speech (Church & Goldin-Meadow, 1986; Goldin-Meadow & Wagner, 2005; Pine et al., 2003). For example, Church and Goldin-Meadow (1986) report that some children who provided incorrect verbal explanations in a Piagetian conservation of liquid task (i.e., emphasizing differences in the height of the glass), produced gestures that referenced the *width* of the glass, indicating their nascent understanding of the relevant variable. Children who produced these speech-inconsistent gestures were more responsive to later instruction than children who produced speech-consistent gestures or no gestures at all. The authors argued that children's spontaneous gestures can sometimes reveal the "thoughts at their fingertips" (Goldin-Meadow & Wagner, 2005), signaling a transitional state during learning and conceptual change. Given that gestures have been proposed to function as a *comfortable middle ground* between concrete and abstract concepts (Novack et al., 2014), gestures produced during the retelling tasks and beyond may provide a novel tool for assessing the emergence of analogical reasoning.

Also, given that we used relatively inclusive criteria for coding gesture content, ongoing work will apply a more fine-grained approach, coding the type and semantic content of gestures, as well as the co-occurrence between spontaneous gestures and speech. We observed a significant difference between children who produced gestures that were *relevant* to the solution performing better at the analogical transfer than children who did not produce any gestures in both Experiment 1 and 2. By analyzing the relative timing of gesture production in relation to online speech production, we would be able to compare whether the temporal co-occurrence with speech (e.g., see examples of comparison of speech and gesture, Iverson & Goldin-Meadow, 1997) or the semantic content itself is more indicative of deeper analogical reasoning. Future work will also examine any differences in the occurrence or type of gestures produced during the first and second retelling tasks. This will allow us to explore specific effects of spontaneous gesture production before and after children had the opportunity to compare solutions between structurally-similar stories (see Cooperrider & Goldin-Meadow, 2014 for related findings in adults). Specifically, the difference in gestures in the two retelling tasks may provide us with insights to teasing apart the two functions of gestures (e.g., Kita et al., 2016), *activating* the abstract concepts (e.g., more likely to happen in the first retelling task) and manipulating the abstract concepts by schematizing them from surface details (e.g., more likely to happen in the second retelling task). In addition, future studies will further explore the relationship between engagement, speech and gesture production, and analogical transfer by including additional measures to capture potential differences in attentiveness (e.g., memory questions). Additional work is also necessary to determine whether spontaneous gestures *facilitated* analogical reasoning or were a by-product of this process.

Conclusion

In sum, Chapter 3 provides initial evidence for the relationship between children's spontaneous gestures and their analogical transfer, suggesting that gesture may provide a window into early abstract reasoning. We find that five- to seven- year-olds who produce gestures when activating relevant knowledge (vs. who do not use gestures) are more likely to later succeed in an analogical transfer task and even more so if children represent solution-relevant gestures in their hands. This chapter contributes to the two fields on early gesture production and analogical reasoning by bridging them with the potential role of gesture production in children's activation and manipulation of abstract concepts in the online process of analogical reasoning.

Acknowledgements

Experiment 1 in Chapter 3 had been published in the proceedings of the 43rd Annual Meeting of the Cognitive Science Society. The dissertation author was the primary investigator and author of this paper.

Thank you to Aarthi Popat, Antonia Gruert, Anika Jallorina, Allison Morgan, Emily Songvilay, Jake Troung, Jolie Nguyen, Julia Hoffman, Kristy Kim, Meryem El-Ansari, Nicky Sullivan, Trisha Katz, and Maddie Weerts. for their assistance with data collection. We thank Kensy Cooperrider for his input on the study design, and Jamie Amemiya and Erik Brockbank for statistical guidance. We are grateful to the Birch Aquarium, Fleet Science Center and all of the participating preschools. Finally, we thank the participants and their parents for making this research happen. This work was supported by a Hellman Foundation Fellowship and a Jacobs Foundation Fellowship to C. Walker.

DISCUSSION

In this dissertation, I presented evidence of the extent to which children readily produce and comprehend meaningful movements with a rich spectrum of abstract goals. Specifically, I focused on unveiling children's ability to produce movement itself as a goal (e.g., early dance to music in Chapter 1), children's ability to flexibly reason about unobserved movements produced by agents (e.g., the self-propelled movement of an agent making music in Chapter 2), and children's ability to produce representational gestures in analogical reasoning (e.g., using gestures to represent structural information when retelling a story in Chapter 3).

In Chapter 1, I characterized early movement to music during the first two years of life based on parents' naturalistic observations of their children. I showed that the motivation and tendency to move to music appears extremely early in life, around 9 months of age. Movement to music also exhibited developmental change over the first two years, showing a quantitative increase in the frequency of dance and qualitative changes in children's mimicry of iconic gestures. Based on these findings, movement to music appears to be a fundamental part of infants' behavioral repertoire, one that is shaped by both maturation and learning.

In Chapter 2, I zoomed out to examine how children reason about others' intentional movement, especially when the movement itself is not directly observed but can be inferred from a combination of contextual cues (e.g., a musical staircase, melodies with different timing of notes). I showed that from age 6 and onward, children can demonstrate an adult-like level of causal reasoning in linking music with agents and movements. When 6-year-olds heard some sounds being played on a musical staircase behind a curtain, they were more likely to infer that an agent produced the music when it would have required self-propelled movements to override the physical constraints of the environment. I also suggested a potential developmental change in causal reasoning about movements and music. Unlike older children, 4-year-olds were biased in attributing all orderly outcomes to agents, regardless of the physical context in which the music was created.

In Chapter 3, I shifted to focus on the specific relationship between analogical reasoning and representational movements produced by children by looking at co-speech gestures. I presented a significant correlation between spontaneous gesture production and analogical transfer success. Children between 5 and 7 years of age who spontaneously produced gestures during a communicative task (e.g., retelling a story) performed better later at an analogical transfer task than those who did not produce gestures. I also found support for the importance of self-generated gestures with evidence that children who were asked to sit on their hands seemed to have a harder time drawing analogies across different contexts.

The three chapters each contribute two meaningful implications to the field: highlighting the value of unprompted naturalistic observations in furthering the next steps in the literature and expanding the methodological horizon for studies observing non-object-directed goals. First, I demonstrate the importance of revisiting real-life naturalistic observations for fine-tuning observations of child behavior. To my knowledge, Chapter 1 was the first study to look extensively at multiple variables about early dance from well-experienced observers of children, their parents. Whereas previous work on movement to music was based on well-controlled in-lab studies (e.g., Fujii et al., 2014, de l'Etoile et al., 2020), participating children in these studies were asked to behave naturally in an unfamiliar setting, which may have suppressed their behavior. Leveraging parents' rich and consistent observation of children's dance behavior, I was able to capture both qualitative and quantitative developmental changes within each individual child and across infants as a group. Such valuable observations were possible in naturalistic settings by familiar observers. On a similar note, Chapter 3 was the first study to examine children's spontaneous gestures, ones that were not explicitly taught, as a window into analogical reasoning. Several studies in early gesture production have focused on manipulating the use of gestures by explicitly training children to use gestures and teaching them which gestures to use in a controlled manner (e.g., Novack et al., 2016; Wakefield et al., 2019). Although these studies have discovered strong relationships between gesture production and manipulation of abstract concepts, they are not able to capture the naturalistic use of gestures by children in unprompted contexts or measure what *types* of gestures would be spontaneously used by children if they were to do the same task on their own. In contrast, the first experiment in Chapter 3 enabled researchers to observe both the frequency of spontaneous gestures and the format in which abstract concepts were represented by children's "fingertips."

Second, these studies expand the horizon of developmental research methods by providing concrete examples of closely capturing the construct of non-object-directed goals. In Chapter 1, I used an operational definition that is more tangible and imaginable to parents. Previously, measures of children's movement to music had been heavily constrained by definitions of rhythmic entrainment, which required being on the beat most of the time when moving to music (de l'Etoile et al., 2020; Fujii et al., 2014). I provided a definition of dance as something broader, indicating that it doesn't have to be perfectly aligned with the beat all the time but is something that happens contingent on the presence of music. By providing caregivers with a big-picture perspective that validated their observations, I enabled them to share a rich spectrum of scientific observations on the early development of dance. In Chapter 2, I provided adults and children with tools that helped them expand on the hypothetical space of music-making for causal reasoning. Even as adults, we might default to reason that music is produced by animate agents (e.g., a singer, a violinist) rather than inanimate objects (e.g., the wind going through wind chimes and making sounds). Therefore, I introduced two tangible/imaginary figures representing each end of the spectrum of agency: an inanimate object (e.g., a ball) and an animate agent (e.g., Fred). By triggering the spectrum of potential causes of music, I was able to discover compelling evidence for a developmental trajectory in early causal reasoning about music and agency in childhood from age 4 to age 6. In Chapter 3, I designed studies to capture the "online" processing of children's causal reasoning by observing co-speech gestures along with speech. By enabling children to freely retell a story in Experiment 1, I was able to achieve two goals: observing *when* gestures were recruited in alignment with the content of speech and minimizing the *cognitive load* of this process. I presented evidence that co-speech gestures used in retellings were something beyond the "tip-of-the-tongue" phenomenon. That is, the co-speech gestures specifically occurred when the children's speech addressed the structural information of the different stories, potentially increasing the rate of analogical transfer. Such observations would not have been possible in previous studies that trained children to use gestures as a "means" to achieve another goal (e.g., Wakefield et al., 2019). Therefore, this dissertation demonstrates thinking outside the box to discover the usage and understanding of meaningful movements in childhood, building stepping stones for future research.

Altogether, these studies theoretically contribute to the literature by providing additional windows to reasoning and production of meaningful movements. Chapter 1 and Chapter 2 together provides a rich characterization about the development of musicality through non-object directed movements. As Schachner & Carey (2013) suggested, goals that are not necessarily object-directed may still signal abstract goals such as the movement itself being a goal (e.g., movement-based goals). Chapter 1 validates that children do *produce* these movement-based goals and does so frequently and earlier than what was observed in previous literature. Chapter 2 presented how flexibly adults and older children can *reason* about the movements involved in music-making thereby complementing our understanding of development of musicality. Altogether, Chapter 1 and 2 suggests that music, which has rich structural information but also is unobservable by sight, can provide an additional window into how music is perceived and what is reasoned from music beyond visually accessing the goal.

Chapter 1 and Chapter 3 both suggest that children's self-*produced* movements that does not involve objects still serves unique goals at their "finger tips": movement-based goals (Schachner & Carey, 2013) and representational gestures that activate relevant knowledge (Kita et al., 2017; Novack et al., 2016). Chapter 1 presents that children readily and frequently produce movement-based goals and that its format (e.g., the use of iconic gestures, the use of different movements) becomes diversified with age. Potential future studies may expand on looking at the intrinsic motivation for producing movement-based goals with time and effort and how these movements contribute to children's emotional and/or social development (e.g., What does it mean to enjoy dancing to music?). Chapter 3 suggests that the representational gestures produced by children may signal the activation and/or manipulation of relevant concepts for problem solving (Kita et al., 2017). Specifically, I find that the more semantically relevant content was gestured the analogical transfer success was higher than those who did not gesture at all. Chapter 1 and Chapter 3 both suggests that observing children's movements produced, ideally in its naturalistic form (e.g., at home in Chapter 1; spontaneous gestures in Experiment 1 in Chapter 3), can signal children's deeper engagement with abstract goals, beyond those represented as objects.

I also want to acknowledge limitations in the three studies that could be addressed by follow-up studies in the future. First, the data collection method might need more replications to make it generalizable to a representative sample. Chapter 1 drew from parental reports that were mainly based on voluntary participation in a questionnaire, but such a sample might have been biased to include parents who were already actively observing their children or who had more flexible hours that they could devote to something beyond their daily routine. Therefore, future studies may aim to replicate the present study with a more representational sample that could address different demographic factors, such as race and ethnicity and socio-economic status. Also, the parent questionnaire was mainly answered by parents who live in the United States, but different cultures may have different norms for or access to music and dance. Future studies may expand the sample to include families in different cultures to observe cross-cultural differences, especially to look at what aspects of culture may shape children's early movement to music. In terms of replicability, Chapter 2 includes participant data from both the pre-pandemic and pandemic eras, from 2018 to 2022. Therefore, a good portion of in-person participants in Experiment 2 had to wear masks during the task and also heard the instructions from experimenters in masks. I observed 5- to 7-year-old children becoming more shy or reluctant about telling a story to an unknown child during the pandemic era, whereas those in the pre-pandemic era could routinely interact with multiple unknown children at a nearby playground without bashfulness. Future studies may be able to replicate Experiment 2 in a setting that more closely resembles a normal daily scene for children, perhaps without mandates on mask-wearing.

Second, the task difficulty and measures for uncertainty were not as well adjusted for Chapter 2 as would have been optimal. Although 6-year-old children demonstrated the same level of flexible causal reasoning in Experiment 1b as adults, children in the same age range in Experiment 2b did not necessarily demonstrate the same level of causal reasoning when the dependent measure was changed. It could be the case that answering a question about the cause of music (e.g., "What came out of the pipe?" in Experiment 1b) when knowing the physical environment behind the curtain was perceptually easier than imagining the physical context while knowing who it was that was making the music (e.g., "Which xylophone was behind the curtain?" in Experiment 2b). Still, such a gap in 6-year-olds' performance may have been due to presenting a middle option (e.g., "could be either") in addition to the 2 alternative forced choice questions (2AFCs). Six-year-olds might have misinterpreted this option as a representation of uncertainty (e.g., "I do not know."), rather than representing both possibilities of the answer (e.g., "Both choices can be answers to this question."). Future studies can examine the fine line between uncertainty and representation of multiple possibilities by scaffolding children with the dependent measure.

Third, the manipulation in Chapter 3 was not as strong as expected. We did not find significant differences between the two manipulations (e.g., promoting gesture usage by having children watch someone else gesture vs. having them sit on their hands to constrain their usage of gestures). Rather, the condition promoting the usage of gesture showed a similar pattern of analogical transfer to that of the condition that did not prompt spontaneous gestures. Therefore, follow-up studies may add new conditions that address different combinations of variables. For example, to tease apart the mere effect of exposure to the experimenter's retelling (speech) from the effect of having access to hands, a condition could be added in which the child hears the story retold from an experimenter who does not use co-speech gestures. Also, a condition in which the variables potentially seem to conflict may help future researchers revisit the importance of self-produced gestures. For example, a condition in which the child watches the experimenter's retelling with co-speech gestures but is asked to sit on their hands could be compared to a condition in which the child is able to use their hands but not see the experimenter's retelling with gestures.

Conclusion

Overall, this dissertation presents three insightful observations about children's reasoning and production of meaningful movements. By observing children's self-generated movement to music, I discovered a distinctive area of developmental repertoire. By observing children's reasoning about music and movements, I found an intriguing developmental trajectory of potential biases in understanding social agents that get fine-tuned with experience. By observing children's co-speech gestures "online" as they process structural information, it is possible to gauge how deeply they are engaging in drawing analogies. The three studies address the importance of expanding our scope beyond current behavioral methods in quantifying movements in children's life and revisiting the naturalistic context to discover unexplained variances that may play a crucial role in early social, musical, and cognitive development.

REFERENCES

Agrawal, T., & Schachner, A. (2022). Hearing water temperature: Characterizing the development of nuanced perception of sound sources. *Developmental Science*. https://doi.org/10.1111/desc.13321.

Alibali, M. W., Kita, S., & Young, A. (2000). Gesture and the process of speech production: We think, therefore we gesture. *Language and Cognitive Processes, 15*, 593–613.

Bachner-Melman, R., Dina, C., Zohar, A. H., Constantini, N., Lerer, E., Hoch, S., Sella, S., Nemanov, L., Gritsenko, I., Lichtenberg, P., Granot, R., & Ebstein, R. P. (2005). AVPR1a and SLC6A4 gene polymorphisms are associated with creative dance performance. *PLoS Genetics*, *1*(3), e42. https://doi.org/10.1371/journal.pgen.1008135.

Bahrick, L. E. (1988). Intermodal learning in infancy: Learning on the basis of two kinds of invariant relations in audible and visible events. *Child Development*, *59*(1),197–209.

Bahrick, L. E. (1983). Infants' perception of substance and temporal synchrony in multimodal events. *Infant Behavior and Development, 6*(4), 429–451.

Bainbridge, C.M., Bertolo, M., Youngers, J., Atwood, S, Yurdum, L., Simson, J., Lopez, K., Xing, F., Martin, A. & Mehr, S.A. (2021). Infants relax in response to unfamiliar foreign lullabies. *Nature Human Behavior, 5*, 256-264. https://doi.org/10.1038/s41562-020-00963-z

Baker, C. L., Jara-Ettinger, J., Saxe, R., & Tenenbaum, J. B. (2017). Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behaviour*, *1*(4), 1-10.

Baker, C. L., Saxe, R. R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition, 113*(3), 329–349.

Baruch, C., and Drake, C. (1997). Tempo discrimination in infants. Infant Behavior and Development, 20, 573–577. 10.1016/S0163-6383(97)90049-7

Battaglia, P. W., Hamrick, J. B., & Tenenbaum, J. B. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences*, *110*(45), 18327–18332.

Benetti, L., & Costa-Giomi, E. (2020). Infant vocal imitation of music. *Journal of Research in Music Education*, *67*(4), 381-398. https://doi.org/10.1177/0022429419890328

Blust, A. S., Baker, D. J., Richard, K., & Shanahan, D. (2016). Music, animacy, and rubato: what makes music sound human?. In *Proceedings of the 14th international Conference on Music Perception and Cognition* (pp.186-190). CA: San Francisco.

Brand, R. J. & Shallcross, W. L. (2008). Infants prefer motionese to adult-directed action. *Developmental Science, 11*, 853-861. 10.1111/j.1467-7687.2008.00734.x

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

Brown, A. L., Kane, M. J., & Echols, C. H. (1986). Young children's mental models determine analogical transfer across problems with a common goal structure. *Cognitive Development, 1*(2), 103-121.

Brown, S. E. (1991). Determination of location of pitch within a musical vibrato. *Bulletin of the Council for Research in Music education, 108,* 15-30. Stable URL: https://www.jstor.org/stable/40318435

Cannon, J. J., & Patel, A. D. (2021). How Beat Perception Co-opts Motor Neurophysiology. *Trends in Cognitive Sciences, 25*(2), 137–150. https://doi.org/10.1016/j.tics.2020.11.002

Chen-Hafteck, L., & Mang, E. (2012). Music and language in early childhood development and learning. *The Oxford Handbook of Music Education*, *1*, 261-278.

Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008). Listening to musical rhythms recruits motor regions of the brain. *Cerebral Cortex, 18*(12), 2844–2854.

Christie, S., & Gentner, D. (2010). Where hypotheses come from: Learning new relations by structural alignment. *Journal of Cognition and Development, 11*(3), 356-373.

Church, R. B., & Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition, 23,* 43-71.

Church, R. B., Alibali, M. W., & Kelly, S. D. (Eds.). (2017). *Why Gesture?: How the hands function in speaking, thinking and communicating* (Vol. 7). John Benjamins Publishing Company.

Cirelli, L. K., Einarson, K. M., & Trainor, L. J. (2014). Interpersonal synchrony increases prosocial behavior in infants. *Developmental Science, 17*(6): 1003–11. https://doi.org/10.1111/desc.12193

Cirelli, L. K., Trehub, S. E., & Trainor, L. J. (2018). Rhythm and melody as social signals for infants. *Annals of the New York Academy of Sciences, 1423*, 66-72. https://doi.org/10.1111/nyas.13580

Cirelli, L. K., & Trehub, S. E. (2019). Dancing to Metallica and Dora: Case study of a 19-month-old. *Frontiers in Psychology, 10*, 1073. https://doi.org/10.3389/fpsyg.2019.01073

Conard, N., Malina, M. & Münzel, S. (2009). New flutes document the earliest musical tradition in southwestern Germany. *Nature 460,* 737–740. https://doi.org/10.1038/nature08169

Cooper, R., & Aslin, R. (1990). Preference for infant-directed speech in the first month after birth. *Child Development, 61*(5): 1584-95. https://doi.org/10.1111/j.1467-8624.1990.tb02885.x

Cooperrider, K., & Goldin-Meadow, S. (2014). The role of gesture in analogical problem solving. *In Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 36, No. 36)*.

Cooperrider, K., Gentner, D., & Goldin-Meadow, S. (2016). Spatial analogies pervade complex relational reasoning: Evidence from spontaneous gestures. *Cognitive Research: Principles and Implications, 1*(1), 1-17.

Corbeil, M., Trehub, S. E., & Peretz, I. (2016). Singing delays the onset of infant distress. *Infancy, 21*(3), 373-391. https://doi.org/10.1111/infa.12114

Corrigall, K. A., & Trainor, L. J. (2014). Enculturation to musical pitch structure in young children: evidence from behavioral and electrophysiological methods. *Developmental Science, 17*(1), 142–158. https://doi.org/10.1111/desc.12100

Creighton, D. E., & Sauve, R. S. (1988). The Minnesota Infant Development Inventory in the developmental screening of high-risk infants at eight months. *Canadian Journal of Behavioural Science/Revue Canadienne des Sciences du Comportement*, *20*(4), 424.

Cusimano, M., Hewitt, L. B., Tenenbaum, J., McDermott, J. H. (2018). Auditory scene analysis as Bayesian inference in sound source models. In C. Kalisch, M. Rau, T. Rogers, & J. Zhu (Eds.), *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (pp. 510-516). Madison, WI: Cognitive Science Society.

de l'Etoile, S. K., Bennett, C., & Zopluoglu, C. (2020). Infant Movement Response to Auditory Rhythm. *Perceptual and Motor Skills, 127*(4), 651–670. https://doi.org/10.1177/0031512520922642

Demany, L., McKenzie, B., and Vurpillot, E. (1977). Rhythm perception in early infancy. *Nature 266*, 718–719. 10.1038/266718a0

Drake, C., Jones, M. R., & Baruch, C. (2000). The development of rhythmic attending in auditory sequences: attunement, referent period, focal attending. *Cognition, 77*(3), 251-288. https://doi.org/10.1016/S0010-0277(00)00106-2

Dunn, R.E. (1997). Creative Thinking and Music Listening. *Research Studies in Music Education* 8, 42–45.

Eaves Jr, B. S., Feldman, N. H., Griffiths, T. L., & Shafto, P. (2016). Infant-directed speech is consistent with teaching. *Psychological review*, *123*(6), 758. 10.1037/rev0000031

Fernald, A. (1992). Human maternal vocalizations to infants as biologically relevant signals: An evolutionary perspective. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 391–428). Oxford University Press.

Fritzley, V. H,, & Lee, K. (2003). Do young children always say yes to yes–no questions? A metadevelopmental study of the affirmation bias. *Child development, 74*(5), 1297-1313.

Fujii, S., Watanabe, H., Oohashi, H., Hirashima, M., & Nozaki, D., Taga, G. (2014). Precursors of Dancing and Singing to Music in Three- to Four-Months-Old Infants. *PLoS ONE 9*(5): e97680. 10.1371/journal.pone.0097680

Gaver, W. W. (1993). What in the World Do We Hear?: An Ecological Approach to Auditory Event Perception. *Ecological Psychology, 5*(1), 1–29.

Gelman, R., Durgin, F., & Kaufman, L. (1995). Distinguishing between animates and inanimates: Not by motion alone. In D. Sperber, D. Premack, & A. J. Premack (Eds.), Symposia of the Fyssen Foundation. *Causal cognition: A multidisciplinary debate* (p. 150–184). Clarendon Press/Oxford University Press.

Gentner, D. (1988). Metaphor as structure mapping: The relational shift. *Child development, 47*-59.

Gentner, D., & Hoyos, C. (2017). Analogy and abstraction. *Topics in Cognitive Science, 9*(3), 672-693.

Gentner, D., & Rattermann, M. J. (1991). Language and the career of similarity. *Center for the Study of Reading Technical Report, 533*.Gentner, D., & Toupin, C. (1986). Systematicity and surface similarity in the development of analogy. *Cognitive science, 10*(3), 277-300.

Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, *56*, 165–193.

Gerstenberg, T., Siegel, M. H., Tenenbaum, J. B. (2018) What happened? Reconstructing the Past From Vision and Sound. In C. Kalisch, M. Rau, T. Rogers, & J. Zhu (Eds.), In *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (p.409), Madison, WI.

Ghazban, N. (2013). Emotion regulation in infants using maternal singing and speech. *PhD dissertation*, Ryerson University, Toronto.

Ghetti, S., Hembacher, E., & Coughlin, C. A. (2013). Feeling uncertain and acting on it during the preschool years: A metacognitive approach. Child Development Perspectives, https://doi.org/10.1111/cdep.12035.

Gingras, B., Honing, H., Peretz, I., Trainor, L. J., & Fisher, S. E. (2015) Defining the biological bases of individual differences in musicality. *Philosophical Transactions of Royal Society, 370*: 20140092. http://dx.doi.org/10.1098/rstb.2014.0092

Glascoe, F. P., & Dworkin, P. H. (1995). The role of parents in the detection of developmental and behavioral problems. *Pediatrics,* 95(6), 829-836. https://doi.org/10.1542/peds.95.6.829

Goldin-Meadow, S., & Wagner, S. M. (2005). How our hands help us learn. *Trends in Cognitive Sciences, 9,* 234-241.

Gopnik, A., & Sobel, D.M. (2000). Detecting Blickets: How young children use information about novel causal powers in categorization and induction. *Child Development*, *71*(5), 1205-1222.

Gottfried, A. W., Guerin, D., Spencer, J. E., & Meyer, C. (1984). Validity of Minnesota Child Development Inventory in screening young children's developmental status. *Journal of Pediatric Psychology*, *9*(2), 219-23. https://doi.org/10.1093/jpepsy/9.2.219

Grahn, J. A., & Rowe, J. B. (2009). Feeling the beat: premotor and striatal interactions in musicians and nonmusicians during beat perception. *Journal of Neuroscience*, *29*(23), 7540-7548.

Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological Review, 116*(4), 661–716. doi:10.1037/a0017201

Groarke, J. M., & Hogan, M. J. (2016). Enhancing wellbeing: An emerging model of the adaptive functions of music listening. *Psychology of Music, 44*(4), 769-791.

Guarino, K. F., & Wakefield, E. M. (2020). Teaching analogical reasoning with co-speech gesture shows children where to look, but only boosts learning for some. *Frontiers in Psychology, 11*, 575628.

Guarino, K. F., Wakefield, E. M., Morrison, R. G., & Richland, L. E. (2021). Exploring how visual attention, inhibitory control, and co-speech gesture instruction contribute to children's analogical reasoning ability. *Cognitive Development, 58*, 101040.

Gudmundsdottir, H., & Trehub, S. (2018). Adults recognize toddlers' song renditions. *Psychology of Music, 46*(2), 281-291. https://doi.org/10.1177/0305735617711762

Guerin, D., & Gottfried, A. W. (1987). Minnesota Child Development Inventories: Predictors of intelligence, achievement, and adaptability. *Journal of Pediatric Psychology*, *12*(4), 595-609. https://doi.org/10.1093/jpepsy/12.4.595

Hanna, J. L. (1987). *To Dance is Human: A Theory of Nonverbal communication.* Chicago: University of Chicago Press.

Hannon, E. E., & Trainor, L. J. (2007). Music acquisition: effects of enculturation and formal training on development. *Trends in Cognitive Sciences, 11*(11), 466–472. https://doi.org/10.1016/j.tics.2007.08.008

Hannon, E., Schachner, A., & Nave-Blodgett, J. (2017). Babies know bad dancing when they see it: Older but not younger infants discriminate between synchronous and asynchronous audiovisual musical displays. *Journal of Experimental Child Psychology, 159*, 159-174. https://doi.org/10.1016/j.jecp.2017.01.006

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world?. The *Behavioral and Brain Sciences, 33*(2-3), 61–135. https://doi.org/10.1017/S0140525X0999152X

Hoeschele, M., Merchant, H., Kikuchi, Y., Hattori, Y., & ten Cate, C. (2015). Searching for the origins of musicality across species. *Philosophical Transactions of the Royal Society B: Biological Sciences, 370*(1664), 20140094. https://doi.org/10.1098/rstb.2014.0094

Holyoak, K. J., Junn, E. N., & Billman, D. O. (1984). Development of analogical problem-solving skill. *Child development*, *55*(6), 2042-2055.

Honing, H. (2018). Musicality as an upbeat to music: Introduction and research agenda. In H. Honing (Ed.), *The Origins of Musicality* (pp. 3-20). MIT Press.

Honing, H. (2019). *The evolving animal orchestra: in search of what makes us musical*. MIT Press.

Hostetter, A. B., Wieth, M. B., Foster, K., Moreno, K., & Washington, J. (2016). Effects of Gesture on Analogical Problem Solving: When the Hands Lead you Astray. *In 38th Proceedings of Cognitive Science Society.*

Ireton, H. (1992). *Child Development Inventories*. Minneapolis: Behavior Science Systems.

Iversen, J. R., & Patel, A. D. (2008). The beat alignment test (BAT): Surveying beat processing abilities in the general population. *In Proceedings of the 10th International Conference on Music Perception and Cognition (ICMPC10).*

Iverson, J. M., & Goldin-Meadow, S. (1997). What's communication got to do with it? Gesture in children blind from birth. *Developmental Psychology, 33,* 453-467.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*(5), 770–814.

Kaplan, P. S., Jung, P. C., Ryther, J. S., & Zarlengo-Strouse, P. (1996). Infant-directed versus adult-directed speech as signals for faces. *Developmental Psychology, 32,* 880–891. https://doi.org/10.1037/0012-1649.32.5.880

Keehn, R. J. J., Iversen, J. R., Schulz, I., & Patel, A. D. (2019). Spontaneity and diversity of movement to music are not uniquely human. *Current Biology, 29*(13), R621-R622. https://doi.org/10.1016/j.cub.2019.05.035

Keil, F. C., & Newman, G. E. (2015). Order, order everywhere, and only an agent to think: The cognitive compulsion to infer intentional agents. *Mind and Language, 30*(2), 117–139.

Kim, M., & Choi, K.-S. (2003). Access to structural similarity in the analogical problem solving of children. *School Psychology International, 24*, 218–231.

Kim, M., & Schachner, A. (2022, May 17). The Origins of Dance: Characterizing the Development of Infants' Earliest Dance Behavior. Retrieved from https://osf.io/8gqsn/

Kirschner, S., & Ilari, B. (2014). Joint drumming in Brazilian and German preschool children: Cultural differences in rhythmic entrainment, but no prosocial effects. *Journal of Cross-Cultural Psychology, 45*(1), 137-166. https://doi.org/10.1177/0022022113493139

Kirschner, S., & Tomasello, M. (2010). Joint music making promotes prosocial behavior in 4-year-old children. *Evolution and Human Behavior 31*(5): 354–64. https://doi.org/10.1016/j.evolhumbehav.2010.04.004

Kirschner, S., Tomasello, M. (2009). Joint drumming: social context facilitates synchronization in preschool children. *Journal of Experimental Child Psychology, 102*, 299–314. https://doi.org/10.1016/j.jecp.2008.07.005

Kita, S., Alibali, M. W., & Chu, M. (2017). How Do Gestures Influence Thinking and Speaking? The Gesture-for-Conceptualization Hypothesis. P*sychological Review*, 124(3), 245-266.

Kopp, F. (2014). Audiovisual temporal fusion in 6-month-old infants. *Developmental Cognitive Neuroscience, 9*, 56–67.

Kragness, H. E., Anderson, L., Chow, E., Schmuckler, M., & Cirelli, L. K. (2022b). Musical groove shapes children's free dancing. *Developmental Science*, e13249. https://doi.org/10.1111/desc.13249

Kragness, H. E., Johnson, E. K., & Cirelli, L. K. (2022c). The song, not the singer: Infants prefer to listen to familiar songs, regardless of singer identity. *Developmental Science*, *25*(1), e13149. https://doi.org/10.1111/desc.1314

Kragness, H. E., Ullah, F., Chan, E., Moses, R., & Cirelli, L. K. (2022a). Tiny dancers: Effects of musical familiarity and tempo on children's free dancing. *Developmental Psychology*. Epub ahead of print. doi: https://doi.org/10.1037/dev0001363

Kratus, J. (2017). Music listening is creative. *Music Educators Journal*, *103*(3), 46-51.
Lange, R. (1975). *The Nature of Dance: An Anthropological Perspective.* London: MacDonald & Evans Ltd.

Launay, J. (2015). Musical sounds, motor resonance and agency detection. *Empirical Musicology Review, 10*(1), 30–40.

Lewkowicz, D. J. (1992). Infants' responsiveness to the auditory and visual attributes of a sounding/moving stimulus. *Perception & Psychophysics, 52*(5), 519–528.

Lewkowicz, D. J. (1994). Limitations on infants' response to rate-based auditory-visual relations. *Developmental Psychology, 30*(6), 880.

Lipton, P. (2003). *Inference to the best explanation*. Routledge.

Little, E. E., Legare, C. H., & Carver, L. J. (2019). Culture, carrying, and communication: Beliefs and behavior associated with babywearing. *Infant Behavior and Development,* 57, 101320. 10.1016/j.infbeh.2019.04.002

Louie, J. Y., Wang, S., Fung, J., & Lau, A. (2014). Children's emotional expressivity and teacher perceptions of social competence. I*nternational Journal of Behavioral Development, 39*(6), 497–507. 10.1177/0165025414548775.

Ma, L., & Xu, F. (2013). Preverbal infants infer intentional agents from the perception of regularity. *Developmental Psychology*, *49*(7), 1330–1337.

Madison, G. (2006). Experiencing groove induced by music: consistency and phenomenology. *Music Perception*, *24*(2), 201-208. https://doi.org/10.1525/mp.2006.24.2.201

Madison, G., Gouyon, F., Ullén, F., & Hörnström, K. (2011). Modeling the tendency for music to induce movement in humans: First correlations with low-level audio descriptors across music genres. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(5), 1578. https://doi.org/10.1037/a0024323

ManyBabies Consortium. (2020). Quantifying sources of variability in infancy research using the infant-directed-speech preference. *Advances in Methods and Practices in Psychological Science*, *3*(1), 24-52. https://doi.org/10.1177/2515245919900809

Margulis, E. H., Wong, P. C., Simchy-Gross, R., & McAuley, J. D. (2019). What the music said: Narrative listening across cultures. *Palgrave Communications, 5*(1), 1-8.

Matsumoto, D., Yoo, S. H., Fontaine, J., Anguas-Wong, A., Arriola, M., & Ataca, B. (2008). Mapping expressive differences around the world: The relationship between emotional display rules and individualism versus collectivism. *Journal of Cross-Cultural Psychology, 39*(1), 55–74. https://doi.org/10.1177/0022022107311854

McAuley, J. D., Jones, M. R., Holub, S., Johnston, H. M., & Miller, N. S. (2006). The time of our lives: life span development of timing and event tracking. *Journal of Experimental Psychology: General*, *135*(3), 348. https://doi.org/10.1037/0096-3445.135.3.348

McKinney, C. H., Tims, F. C., Kumar, A. M., & Kumar, M. (1997). The effect of selected classical music and spontaneous imagery on plasma β-endorphin. *Journal of Behavioral Medicine*, *20*(1), 85-99. DOI: 10.1023/a:1025543330939.

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

Mehr, S. A. (2014). Music in the Home: New Evidence for an Intergenerational Link. *Journal of Research in Music Education, 1*(11). https://doi.org/10.1177/0022429413520008

Mehr, S. A., Krasnow, M. M., Bryant, G. A., & Hagen, E. H. (2021). Origins of music in credible signaling. *Behavioral and Brain Sciences*, *44*, e60:23-39. 10.1017/S0140525X20000345.

Mithen, S. J. (2005). *The singing Neanderthals: The origins of music, language, mind, and body*. London: Weldenfeld & Nicholson.

Müllensiefen, D., Gingras, B., Stewart, L. & Musil, J. (2014). *The Goldsmiths Musical Sophistication Index (Gold-MSI): Technical Report and Documentation v1.0*. London: Goldsmiths, University of London.

Nakata, T., & Trehub, S. E. (2004). Infants' responsiveness to maternal speech and singing. *Infant Behavior and Development, 27*(4): 455–64. https://doi.org/10.1016/j.infbeh.2004.03.002

National Research Council. (2008). *Early childhood assessment: Why, what, and how*. National Academies Press. URI: http://digilib.umpalopo.ac.id:8080/jspui/handle/123456789/388

Nave-Blodgett, J. E., Snyder, J. S., & Hannon, E. E. (2021). Hierarchical beat perception develops throughout childhood and adolescence and is enhanced in those with musical training. *Journal of Experimental Psychology: General, 150*(2), 314. https://doi.org/10.1037/xge0000903

Nettl, B. (2015). *The study of ethnomusicology: Thirty-three discussions (3rd ed.).* Champaign: University of Illinois Press. Stable URL: https://www.jstor.org/stable/44734974

Newman, G. E., Keil, F. C., Kuhlmeier, V. A., & Wynn, K. (2010). Early understandings of the link between agents and order. *Proceedings of the National Academy of Sciences of the United States of America*, 107(40), 17140–17145.

Novack, M. A., Congdon, E. L., Hemani-Lopez, N., & Goldin-Meadow, S. (2014). From action to abstraction: Using the hands to learn math. *Psychological Science, 25*, 903-910.

Oliver, B., Dale, P., Saudino, K., Petrill, S., Pike, A., Plomin, R. (2002). The validity of a parent-based assessment of cognitive abilities in three-year olds. *Early Child Development and Care.* 172(4):337–348. https://doi.org/10.1080/03004430212713

Outa, J., Zhou, X. J., Gweon, H., & Gerstenberg, T. (2022). Stop, children what's that sound? Multi-modal inference through mental simulation. Proceedings of the 44th Annual Cognitive Science Society.

Patel A.D. (2021). Vocal learning as a preadaptation for the evolution of human beat perception and synchronization. *Phil. Trans. R. Soc. B,* 376: 20200326. http://doi.org/10.1098/rstb.2020.0326

Patel, A. D. (2008). *Music, language, and the brain*. New York, NY, US: Oxford University Press.

Patel, A. D., & Iversen, J. R. (2014). The evolutionary neuroscience of musical beat perception: the Action Simulation for Auditory Prediction (ASAP) hypothesis. *Frontiers in Systems Neuroscience*, *8*, 57.

Pearl, J. (2000.) *Causality: Models, Reasoning, and Inference.* Cambridge, UK: Cambridge Univ. Press

Peterson, E.M. (2006). Creativity in Music Listening. *Arts Education Policy Review*, 107 (3), 15–21.

Pine, K. J., Lufkin, N., & Messer, D. (2004). More gestures than answers: Children learning about balance. *Developmental Psychology, 40,* 1059– 1067. http://dx.doi.org/10.1037/0012-1649.40.6.1059.

Politimou, N., Stewart, L., MuÈllensiefen, D., & Franco, F. (2018). Music@Home: A novel instrument to assess the home musical environment in the early years. *PLoS One, 13*(4): e0193819. https://doi.org/10.1371/journal.pone.0193819.

Provasi, J., & Bobin-Bègue, A. (2003). Spontaneous motor tempo and rhythmical synchronisation in 2½-and 4-year-old children. *International Journal of Behavioral Development, 27*(3), 220-231. https://doi.org/10.1080/01650250244000290

Putkinen, V., Tervaniemi, M., & Huotilainen, M. (2013). Informal musical activities are linked to auditory discrimination and attention in 2-3-year-old children: an event-related potential study. *European Journal of Neuroscience, 37*(4): 654-661. https://doi.org/10.1111/ejn.12049 PMID: 23167769.

Reigado, J., & Rodrigues, H. (2018). Vocalizations produced in the second year of life in response to speaking and singing. *Psychology of Music*, *46*(5), 626-637. https://doi.org/10.1177/0305735617719335

Repp, B.H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review 12*, 969–992. https://doi.org/10.3758/BF03206433.

Richert, R. A., & Smith, E. I. (2011). Preschoolers' quarantining of fantasy stories. *Child Development, 82*(4), 1106-1119.

Rock, A. M., Trainor, L. J., & Addison, T. L. (1999). Distinctive messages in infant-directed lullabies and play songs. *Developmental psychology, 35*(2), 527. https://doi.org/10.1037/0012-1649.35.2.527

Trehub, S. E., Becker, J., & Morley, I. (2015). Cross-cultural perspectives on music and musicality. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1664), 20140096. https://doi.org/10.1098/rstb.2014.0096

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

Royce, A.P. (2002). *The Anthropology of Dance.* Hightstown, N.J.: Princeton Book Company.

Savage, P. E., Brown, S., Sakai, E., & Currie, T. E. (2015). Statistical universals reveal the structures and functions of human music. *Proceedings of the National Academy of Sciences of the USA, 112*(29), 8987–8992. https://doi.org/10.1073/pnas.1414495112

Savage, P. E., Loui, P., Tarr, B., Schachner, A., Glowacki, L., Mithen, S., & Fitch, W. T. (2021). Music as a coevolved system for social bonding. *Behavioral and Brain Sciences*, *44*, e59: 1–22. 10.1017/S0140525X20000333

Saxe. R., Tenenbaum. J. B., Carey, S. (2005). Secret Agents: Inferences about hidden causes by 10- and 12-month-old infants. *Psychological Science*, *16*, 995–1001.

Schachner, A. M. (2012). The evolutionary and cognitive basis of the perception and production of dance (Order No. 3543089). Available from ProQuest Dissertations & Theses A&I. (1175324735). Retrieved from https://www.proquest.com/dissertations-theses/evolutionary-cognitive-basis-perception/docview/1175324735/se-2?accountid=14524.

Schachner, A., & Carey, S. (2013). Reasoning about 'irrational' actions: When intentional movements cannot be explained, the movements themselves are seen as the goal. *Cognition, 129*(2), 309-327. https://doi.org/10.1016/j.cognition.2013.07.006.

Schachner, A., & Kim, M. (2018). Entropy, order and agency: The cognitive basis of the link between agents and order. In C. Kalisch, M. Rau, T. Rogers, & J. Zhu (Eds.), *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (pp.2430-2435). Madison, WI.

Schachner, A., Brady, T. F., Pepperberg, I. M., & Hauser, M. D. (2009). Spontaneous motor entrainment to music in multiple vocal mimicking species. *Current Biology, 19*: 831–36. https://doi.org/10.1016/j.cub.2009.03.061

Schneider, R., Yurovsky, D., & Frank, M. C. (2015). Large-scale investigations of variability in children's first words. Proceedings of the 37th Annual Meeting of the Cognitive Science Society, Pasadena, California, July 22–25, 2015 (pp. 2110–2115). Austin, TX: Cognitive Science Society. ISBN: 978-0-9911967-2-2.

Seifer, R., Sameroff, A., Dickstein, S., Schiller, M., & Hayden, L. C. (2004). Your own children are special: clues to the sources of reporting bias in temperament assessments. *Infant Behavior and Development, 27*(3), 323–341. 10.1016/j.infbeh.2003.12.005.

Shenfield, T., Trehub, S. E., & Nakata, T. (2003). Maternal singing modulates infant arousal. *Psychology of Music, 31*(4), 365-375. https://doi.org/10.1177/03057356030314002

Sievers, B., Polansky, L., Casey, M., & Wheatley, T. (2013). Music and movement share a dynamic structure that supports universal expressions of emotion. *Proceedings of the National Academy of Sciences*, *110*(1), 70-75. https://doi.org/10.1073/pnas.1209023110

Sievers, B., Polansky, L., Casey, M., & Wheatley, T. (2013). Music and movement share a dynamic structure that supports universal expressions of emotion. *Proceedings of the National Academy of Sciences of the United States of America, 110*(1), 70–75.

Spelke, E. S. (1979). Perceiving bimodally specified events in infancy. *Developmental Psychology, 15*(6), 626–636.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science, 10*(1), 89–96.
Stefano, G. B., Zhu, W., Cadet, P., Salamon, E., & Mantione, K. J. (2004). Music alters constitutively expressed opiate and cytokine processes in listeners. *Medical Science Monitor*, *10*(6), MS18-MS27. Full-text PDF: http://www.MedSciMonit.com/pub/vol_10/no_6/5351.pdf

Steinbeis, N., & Koelsch, S. (2011). Affective priming effects of musical sounds on the processing of word meaning. *Journal of Cognitive Neuroscience, 23*(3), 604–621.

Steinberg, S., Liu, T., & Lense, M. D. (2021). Musical engagement and parent-child attachment in families with young children during the Covid-19 pandemic. *Frontiers in Psychology, 12,* 1-10. https://doi.org/10.3389/fpsyg.2021.641733

Sturner, R. A., Funk, S. G., Thomas, P. D., & Green, J. A. (1982). An adaptation of the Minnesota Child Development Inventory for preschool developmental screening. *Journal of Pediatric Psychology*, *7*(3), 295-306. https://doi.org/10.1093/jpepsy/7.3.295

Sullivan, J., Mei, M., Perfors, A., Wojcik, E. H., & Frank, M. C. (2020). SAYCam: A large, longitudinal audiovisual dataset recorded from the infant's perspective. *Open Mind*, 1-10. https://doi.org/10.31234/osf.io/fy8zx

Teglas, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J. B., & Bonatti, L. L. (2011). Pure Reasoning in 12- Month-Old Infants as Probabilistic Inference. *Science, 332*(6033), 1054–1059.

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences*, 10, 309–318.

Thelen, E. (1979). Rhythmical stereotypies in normal human infants. Anim. Behav. 27, 699–715. 10.1016/0003-3472(79)90006-x

Thelen, E. (1981). Rhythmical behavior in infancy: An ethological perspective. Developmental Psychology, 17(3), 237–257. https://doi.org/10.1037/0012-1649.17.3.237

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, *7*(1), 53-71. 10.1207/s15327078in0701_5

Traer, J., Norman-Haignere, S. V., & McDermott, J. H. (2020). Causal inference in environmental sound recognition. *BioRxiv.*

Trainor L. J., Austin C. M., Desjardins R. N. (2000). Is infant-directed speech prosody a result of the vocal expression of emotion? *Psychological Science, 11,* 188–195 10.1111/1467-9280.00240.

Trainor, L. J., Clark, E. D., Huntley, A., & Adams, B. A. (1997). The acoustic basis of preferences for infant-directed singing. *Infant Behavior and Development*, *20*(3), 383-396. https://doi.org/10.1016/S0163-6383(97)90009-6

Trehub, S. E. (2003). The developmental origins of musicality. *Nature neuroscience, 6*(7), 669. https://doi.org/10.1038/nn1084

Trehub, S. E. (2019). Multimodal music in infancy and early childhood. *The Oxford Handbook of Music and the Body*, 383.

Trehub, S. E., & Nakata, T. (2001). Emotion and music in infancy. *Musicae Scientiae, 5(1_suppl),* 37–61. https://doi.org/10.1177/10298649020050S103

Trehub, S. E., & Russo, F. A. (2020). Infant-Directed Singing from a Dynamic Multimodal Perspective: Evolutionary Origins, Cross-Cultural Variation, and Relation to Infant-Directed Speech. In *The Routledge Companion to Interdisciplinary Studies in Singing, Volume I: Development* (pp. 249-261).

Trehub, S. E., Plantinga, J., & Russo, F. A. (2016). Maternal vocal interactions with infants: Reciprocal visual influences. *Social Development, 25*(3), 665-683. https://doi.org/10.1111/sode.12164

Trehub, S. E., Unyk, A. M., Kamenetsky, S. B., Hill, D. S., Trainor, L. J., Henderson, J. L., & Saraza, M. (1997). Mothers' and fathers' singing to infants. *Developmental psychology, 33*(3), 500. https://doi.org/10.1037/0012-1649.33.3.500

Trehub, S.E. (2011). Music lessons from infants. In S. Hallam, I. Cross & M. Thaut (Eds.), *Oxford Handbook of Music Psychology*, Oxford University Press: Oxford.

Trevarthen, C. (1999). Musicality and the intrinsic motive pulse: evidence from human psychobiology and infant communication. *Musicae scientiae, 3(1_suppl),* 155-215. https://doi.org/10.1177/10298649000030S109

Tunçgenç, B., Cohen, E., & Fawcett, C. (2015). Rock with me: The role of movement synchrony in infants' social and nonsocial choices. *Child Development, 86*(3), 976-984. https://doi.org/10.1111/cdev.12354

Tversky, A.; Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*. 185 (4157): 1124–1131. 10.1126/science.185.4157.1124.

U.S. Census Bureau (2020). Census regions and divisions of the United States. *Retrieved from https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf*.

Vlismas, W., Malloch, S., & Burnham, D. (2013). The effects of music and movement on mother-infant interactions. *Early Child Development and Care, 183*(11): 1669–88. https://doi.org/10.1080/03004430.2012.746968.

Wakefield, E. M., Foley, A. E., Ping, R., Villarreal, J. N., Goldin-Meadow, S., & Levine, S. C. (2019). Breaking down gesture and action in mental rotation: Understanding the components of movement that promote learning. *Developmental Psychology, 55*. 981-993.

Walker, C. M., & Lombrozo, T. (2017). Explaining the moral of the story. *Cognition, 167,* 266-281.

Wallin, N. L., Merker, B., & Brown, S. (Eds.). (2001). *The origins of music*. MIT press.

Welch, G. F. (2006). *The child as musician: A handbook of musical development*, 311-329.

Winkler, I., Háden, G. P., Ladinig, O., Sziller, I., & Honing, H. (2009). Newborn infants detect the beat in music. *Proceedings of the National Academy of Sciences, 106*(7), 2468-2471. https://doi.org/10.1073/pnas.0809035106

Yu, L., & Myowa, M. (2021). The early development of tempo adjustment and synchronization during joint drumming: A study of 18- to 42-month-old children. *Infancy, 26(*4), 635-646. 10.1111/infa.12403

Zentner, M., & Eerola, T. (2010). Rhythmic engagement with music in infancy. *Proceedings of the National Academy of Sciences, 107*(13), 5768-5773. https://doi.org/10.1073/pnas.1000121107