How Pedagogy, Action Expectations, and Executive Function Jointly Shape 3- to 7-Year-Olds' Event Processing and Exploratory Play

by

Diana J. D. Daschel

A dissertation accepted and approved in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

in Psychology

Dissertation Committee:

Dare A. Baldwin, Chair

Sara D. Hodges, Core Member

Louis J. Moses, Core Member

Melissa A. Redford, Institutional Representative

University of Oregon

Spring 2026

© 2026 Diana J. D. Daschel
This work is openly licensed via CC BY-NC-ND 4.0



DISSERTATION ABSTRACT

Diana J. D. Daschel
Doctor of Philosophy in Psychology

Title: How Pedagogy, Action Expectations, and Executive Function Jointly Shape 3-to 7-Year-Olds' Event Processing and Exploratory Play

Children are immersed in a dynamic and multi-faceted world about which they have much to learn. Several separate bodies of prior research clarify various sources of influence on children's learning but rarely investigate how children integrate such sources during knowledge updating. In this dissertation, I examined ways in which three kinds of factors -- informant characteristics, children's expectations about what they will observe, and individual differences -- might simultaneously and jointly shape their causal learning. I probed these issues in terms of both a) overt exploratory behaviors children displayed in relation to what they'd experienced, and b) children's attentional patterns as the learning events unfolded in time.

Across two studies, children heard either a knowledgeable ("This is how my toy works") or naïve ("I wonder how this toy works") verbal pedagogy statement from an adult informant. The informant then demonstrated a causal action on the toy: either turning on a light in a manner consistent with children's expectations (pressing a button) or violating their expectations (waving above the button). I measured children's attentional patterns via the dwell-time paradigm as they viewed the expectation-consistent or expectation-violating event. Following this, I recorded children's exploratory activities as they played with the light. Lastly, I measured children's executive skill via the Head, Toes, Knees, and Shoulders (HTKS) task.

Regarding exploratory play, children displayed a strong expectation that pressing would activate the toy, yet they also updated this belief based on what the informant said and did. The combination of naïve pedagogy ("I wonder how this toy works") and expectation-violating

action (waving) produced enhanced exploration relative to all other conditions, showcasing that children integrated multiple sources of influence as they actively engaged in knowledge updating. As well, children's executive skill influenced their knowledge updating: stronger executive skills were associated with higher fidelity play and flexibly enhanced exploratory activity.

Children's attentional patterns showed greater causally-relevant systematicity when the event they were viewing accorded with their expectations (press) than when their expectations were violated (wave). Further, a comparison of attentional patterns on second, relative to first, viewing of the event revealed attentional reorganization when naïve pedagogy set the context, regardless of whether the action seen was expectation-consistent (press) or violating (wave). While executive function (EF) did not influence children's attention allocation on its own, it did show relations to exploratory play behaviors in association with children's event-viewing: enhanced attention to causally-relevant event regions was related to greater exploratory behavior dependent upon EF skills.

All in all, the dissertation findings shed light on how children cope with complexity that is ubiquitous in everyday causal-learning scenarios. Multiple sources of influence – pedagogical context, causal expectations, and children's own skill at regulating attention and action – jointly shape what children attend to and how they subsequently act on the world. Each further action children subsequently undertake generates new evidence to spur further inferential updating, an ongoing cognitive decalage into which we are gaining new insight.

ACKNOWLEDGMENTS

I first want to thank my committee members Drs. Louis Moses, Sara Hodges, and Melissa Redford for their feedback, guidance, and support from the initial planning stages of this project to its completion. I am especially grateful for the guidance of Dr. Dare Baldwin, who has the unique capacity to act as both an encouraging advisor, and a brilliant academic mind with an enormous capacity to link the progression of developmental evidence into cohesive scientific explanations. I will always be grateful for the mentorship and example that she provided during my graduate training.

Many research assistants in the Acquiring Minds Lab were a vital part of my dissertation work at various stages, and I am especially thankful for their coding efforts and willingness to travel and assist with data collection--in particular, Megan Sebree. I am grateful to lab mates Dr. Netanel Weinstein and Leah Wothe, who provided engaging scientific dialogue and comradery throughout the course of my program. Thank you also to Dr. Jessica Kosie for repeatedly providing pivotal instruction related to dwell time data handling and analysis.

A number of staff at the Oregon Museum of Science and Industry were instrumental in supporting the data collection for this dissertation (in particular, Annie Douglass, Heather Anderson, Emma Koko, and Chris Cardiel). I thank them for their time and efforts providing detailed feedback on our research plans, hosting multiple training sessions, and coordinating our use of OMSI's Science Playground. Similarly, I thank the staff at the Eugene Science Center (especially Alex Lombardi) for their coordination efforts. I also sincerely thank families from the Portland and Eugene communities for their participation, without whom none of this would have been possible.

Finally, thank you to my father Dr. Daryll DeWald for being the best scientific role model and supportive dad that I could ever ask for. Most of all, thank you to my husband James, who cooked us dinner every night, provided endless encouragement, and engineered a practical solution to every problem that I faced with care.

DEDICATION

This dissertation is dedicated to the memory of my mother Rebecca Joy DeWald, who taught me to always remain curious and treat every living thing with kindness.

TABLE OF CONTENTS

Chapter Page

I. INTRODUCTION 15

Overview 15

Pedagogical Agents as a Source of Evidence 17

A Developmental Lens on Direct Perception and Exploration 23

Violation of Expectation and the Role of Surprise in Exploration 26

Examining Children's Event-Expectations Through Their Dwell Times 28

Approaching Children's Everyday Learning 29

Conclusion 32

II. STUDY 1: THE ROLE OF VERBAL PEDAGOGY, ACTION EXPECTATIONS, AND EXECUTIVE SKILL IN JOINTLY SHAPING CHILDREN'S EXPLORATORY PLAY 34

Introduction 34

Method 38

Participants 38

Materials 40

Novel Object 40

Verbal Pedagogy Videos 41

Expected Action Slideshows 41

Design 43

Measures 45

The Children's Behavior Questionnaire 45

Head Toes Knees Shoulders Task 45
Procedure 46
Coding and Dependent Measures 49
Play Behaviors 49
Memory Check 50
HTKS 51
Results 52
Preliminary Analyses 54
Confirmatory Analyses 58
Exploratory Analyses 64
Discussion 79
III. STUDY 2: DWELL TIME AS A WINDOW INTO HOW VERBAL PEDAGOGY AND ACTION EXPECTATIONS JOINTLY SHAPE CHILDREN'S EVENT PROCESSING 84
Introduction 84
Method 88
Participants 88
Materials 89
Training Stimuli 89
Replication Stimuli 89
Main Dwell Time Stimuli 90
Pedagogy Videos and Novel Object 90
Design 90

Procedure 90
Coding 93
Dwell-Time Exclusions 93
Additional Coding 94
Boundary Judgements 95
Accounting for Pixel Change 95
Data Preparation 97
Results 99
Preliminary Analyses 99
Confirmatory Analyses 102
Exploratory Analyses 106
Discussion 117
IV. GENERAL DISCUSSION 119
Overview 119
Summary of Key Findings 119
Strengths, Limitations, Remaining Questions 124
Broader Implications 129
Future Directions 132
Conclusion 133
V. APPENDICES 135
APPENDIX A 135
APPENDIX B 137
APPENDIX C 140

APPENDIX D 143

VI. REFERENCES 145

LIST OF FIGURES

Figure Page

2.1. Expected Action Slideshow Images 43

2.2 . Design for Study 1 45

2.3. Procedure for Study 1 49

2.4. Mean Proportion of Pressing and Waving by Condition 56

2.5. Memory Recall Responses 57

2.6. Mean Number of Unique Actions by Condition 63

2.7. Age x EF Density Plot for Waving as First Action 66

2.8. Proportion of Time with Demonstrated Action: EF x Expected Action 68

2.9. Proportion of Time with Demonstrated Action: EF x Condition 69

2.10. Mean Number of Unique Actions: EF x Condition 73

2.11. Proportion of Time with Demonstrated Action: EC x Surgency 76

2.12. Proportion of Time with Demonstrated Action: EC x Surgency x Condition ... 78

3.1. Slide-to-Slide Pixel Change by Expected Action Slideshow 97

3.2. Dwell Times for the Reach-Over-Barrier Event: Lab vs. Museum Sample 101

3.3. Comparison of Pixel Change and Log10 Dwell Times 104

3.4. Mean Log10 Dwell Times by Expected Action and Slide Type 105

3.5. Mean Log10 Dwell Times by Verbal Pedagogy and Slide Type 108

3.6. Comparison of Log10 Dwell Times by Condition and Viewing 111

3.7. Log10 Dwell Times by Slide Type x Unique Actions x EF 115

3.8. Log10 Dwell Times by Any Wave x EF 116

LIST OF TABLES

Table Page

2.1. Parent Education Level 39

2.2. Parent-Reported Subjective Social Standing 39

2.3. Participant Pressing and Waving by Condition and Age 55

2.4. Memory Accuracy by Age and Expected Action Group 58

2.5. Factorial ANOVA for the Proportion with Demonstrated Action 60

2.6. Factorial ANOVA for the Number of Unique Actions 60

2.7. Pairwise Comparisons for 2.5 62

2.8. Pairwise Comparisons for 2.6: Two-Way Interaction 63

2.9. Regression: Condition x EF for Proportional Play 67

2.10. Contrast Comparison for 2.9: Two-Way Interaction 67

2.11. Contrast Comparison for 2.9: Three-Way Interaction Part 1 68

2.12. Contrast Comparison for 2.9: Three-Way Interaction Part 2 69

2.13. Regression: Condition x EF for Proportional Play: Other Action 70

2.14. Contrast Comparison for 2.13: Three-Way Interaction Part 1 70

2.15. Contrast Comparison for 2.13: Three-Way Interaction Part 2 71

2.16. Regression: Condition x EF for Unique Actions 72

2.17. Contrast Comparison for 2.16: Three-Way Interaction Part 1 72

2.18. Contrast Comparison for 2.16: Three-Way Interaction Part 2 72

2.19. Regression: EC x Surgency for Proportional Play 75

2.20. Contrast Comparison for 2.19: Two-Way Interaction 75

2.21. Regression: EC x Surgency x Expected Action for Proportional Play 77

2.22. Contrast Comparison for 2.21: Three-Way Interaction Part 1 77

2.23. Contrast Comparison for 2.21: Three-Way Interaction Part 2 77

3.1. Mixed Effects: Slide Type Controlling for Pixel Change for Dwell Time 103

3.2. Mixed Effects: Condition x Slide Type for Dwell Time 107

3.3. Contrast Comparison for 3.2: Two-Way Interaction 107
3.4. Contrast Comparison for 3.2: Non-significant Two-Way Interaction 107
3.5. Mixed Effects: Condition x Slide Type x Viewing for Dwell Time..... 110
3.6. Contrast Comparison for 3.5: Two-Way Interaction 110
3.7. Mixed Effects: Slide Type x Unique Actions x EF 114
3.8. Contrast Comparison for 3.7: Two-Way Interaction 114
3.9. Contrast Comparison for 3.7: Three-Way Interaction Part 1..... 114
3.10. Mixed Effects: Slide Type x Any Wave x EF 116
3.11. Contrast Comparison for 3.10: Two-Way Interaction 116

CHAPTER I

INTRODUCTION

1.1 Overview

Within developmental science, many theorists have characterized children as active explorers facilitating their own knowledge-acquisition rather than passive observers of their environments. Jean Piaget, for example, argued that children construct knowledge by acting upon the world around them and observing related effects (Piaget, 1954). This view—constructivism—posits that with continued evidence-gathering and the development of information-processing skills, children update their beliefs and construct increasingly reliable knowledge systems (Miller, 2010). Though modern empirical evidence has often challenged Piaget's notions of *when* and *how* the cognitive abilities involved in such processes arise (Carey et al., 2015), many developmentalists continue to acknowledge merits of the constructivist framework. Modern versions of constructivism such as *rational constructivism* or *theory theory* describe the child as a scientist, iteratively updating beliefs about the world in the face of new evidence (Fei, 2019; Gopnik & Meltzoff, 1997; Gopnik & Wellman, 2012).

In this dissertation, I rely especially on a line of thinking motivated by a Bayesian framework, which suggests that learners generate hypotheses about what they will encounter in the world, draw inferences from the data that they sample, and generate posterior likelihoods regarding the content they learn about, taking both their hypotheses and the data into account (Bonawitz & Ullman, 2024). One strength of this approach is that it can apply a framework for how learning operates in real time regardless of the details surrounding the origins of mental representations (i.e., whether they are innate or learned) (Weisberg & Sobel, 2022). This view does not suggest that children apply Bayesian inference in an exacting, algorithmic way or

consciously perform complex computational analysis in order to reach conclusions (Perfors & Navarro, 2012). Rather, the idea is that zeroing in on meaningful likelihoods is a strategy that often shows up in children's learning.

One key aspect of this "Bayesian inference" view of learning, from earliest childhood onward, is the idea that children sample information differently in relation to how evidence is presented. According to this view, children hold different sampling assumptions depending upon characteristics of the learning environment. Put another way, they express learning behaviors that are sensitive to situational cues. For example, if cues within the context lead the learner to believe it is pedagogical (i.e., if an adult speaks directly to a child about a given topic with authority, shows examples or visuals of a phenomenon, and engages the child in reflecting on the content), then a learner is more likely to follow a "pedagogical sampling assumption" as they engage in learning (Schneidman et al., 2016). Utilizing a pedagogical sampling assumption, children sample information strategically and non-exhaustively, given that a child's learning is directed by the pragmatic implications of pedagogical scenarios. Such implications include those related to conversational scenarios more generally, such as the idea that a speaker will provide relevant, informative, true, and clear information (Grice, 1975).

On the other hand, if a pedagogical agent fails to provide relevant and informative evidence, or they otherwise show signs of violating the pragmatic features of a pedagogical context (such as appearing uncertain, off topic, or leaving a situation in the middle of teaching and thus violating the idea that their pedagogy was exhaustive), then a learner might instead use a "default sampling assumption." This type of sampling is associated with broad exploration, similar to how a learner would have sampled if they had not heard the pedagogy at all. Within a particular learning context, both pedagogy-related features (e.g., language utilized, markers of

authority, tone of voice) and direct evidence (i.e., what a learner directly perceives about an object or topic) will influence a child's emerging beliefs. My overarching aim in this dissertation is to better understand the joint influence of pedagogically-conveyed information alongside direct experience as children are in the grip of learning.

More specifically, I aim to investigate a) what young children infer from the pedagogical intentions of an adult and how this shows up in their exploratory play, b) what they directly experience from observing events unfolding and how this shows up in their patterns of attention, and c) how these elements together influence how children sample information. In what follows, I review available evidence concerning how children learn from both adult pedagogy and from directly exploring their environments and processing related events. I will end by discussing how the development of executive function promotes flexibility in children's learning related to these factors. This review will give shape to a set of novel research questions that motivate the studies I have undertaken and the findings they have generated.

1.2 Pedagogical Agents as a Source of Evidence in Children's Knowledge-Acquisition

Epistemological accounts of knowledge tend to partition evidence into foundational sources, such as direct experience via one's sense perception, inductive reasoning, and memory (Chisholm, 1957), and secondary evidence presented by other people from written or oral explanations (Goldberg, 2008), with the latter type sometimes referred to as "testimony" (Harris, et al., 2018). Both types of evidence commonly influence the formation of beliefs, although some epistemologists argue that genuine knowledge can only arise from true and justified beliefs that are based in foundational sources of evidence (Gettier, 1963)—a discussion for another dissertation. Children's evidence-gathering typically arises through some combination of these two kinds of evidential sources, where children both directly perceive a phenomenon and gain

insight about it from pedagogical agents. However, there are certain topics (such as history) that primarily involve evidence presented by others, and trust of others is essential to the efficacy of children's learning across many topics.

Also in the mix are children's direct perceptions of pedagogical agents themselves, which can alter the likelihood that they trust the veracity of evidence presented to them by such agents (and this in turn can change children's sampling assumptions, as discussed above). A large literature on children's selective trust within learning contexts has investigated how 3- to 6-year-olds selectively choose pedagogical agents based upon both the epistemic (e.g., knowledgeability, accuracy) and social (e.g., warmth, appearance, accent) cues that they convey. In general, findings suggest that children are more likely to ask questions and endorse the statements of agents who possess positive social and reliable epistemic characteristics (Harris et al., 2018; Corriveau & Kurkul, 2014), and that this tendency grows stronger with age (Brosseau-Liard et al., 2014). However, when pedagogical cues vary along both social and epistemic dimensions simultaneously (for example, one agent is nice but inaccurate and one is mean but accurate), preschoolers tend to endorse the pedagogy presented by agents who possess reliable epistemic but negative social characteristics, over those who show positive social attributes but low knowledgeability (Tong et al., 2020). This indicates that in pedagogical situations, children weigh multiple cues when deciding whether to trust in the teaching of an agent, yet they tend to especially prioritize an agent's knowledgeability rather than their warmth when choosing whom to trust as a teacher.

It should be noted that the selective trust paradigm does not reflect the realities of children's everyday learning contexts (i.e., children are not typically presented with a choice among sources who vary along a controlled set of characteristics). Further, although preschool

and kindergarten-age children possess the ability to prioritize among pedagogical agents based upon epistemic criteria, they sometimes remain credulous even when directly presented with evidence that conflicts with an agent's teaching (Jaswal et al., 2014). Evidence suggests that children's trust in pedagogy is an adaptive learning strategy that allows them to efficiently reach conclusions, and they can flexibly deploy this trust depending upon the particulars of the learning context, yet they don't always reach accurate conclusions as a result. The selective trust findings merely indicate that with age, children increasingly possess the ability to select pedagogical agents who possess reliable epistemic traits within contexts where they are provided with multiple sources to choose from. Specifically, children prioritize agents' knowledgeability over warmth when deciding whom to learn from.

While the selective trust paradigm allows us to understand more about how children prioritize pedagogical cues in their learning, it does not demonstrate a typical learning encounter. In their everyday lives, children are embedded within environments where they hear claims from a variety of pedagogical agents either one-on-one or in groups. These agents differ greatly in their ostensible claims and their relationships to children. However, children have an overwhelming propensity to consult adults in their search for explanatory evidence (i.e., to ask them questions, observe their actions, and copy their behaviors) (Frazier et al., 2010), and to believe the claims of adults (Jaswal et al., 2010). This accords with a framework called *natural pedagogy*, or the idea that children are generally prepared to learn from adults and that this is a biologically-based tendency arising in infancy (Csibra & Gergely, 2009). Evidence suggests that, beginning in infancy, children infer that speakers are sharing meaningful information as they communicate (Tomasello et al., 2007). As children progress developmentally, they gain ever-more-sophisticated skills for understanding and integrating the information others offer. For

example, preschool-age children show an emerging understanding of others' minds as representational, which develops in concert with their executive function abilities (Perner, 1993; Moses et al., 2005; Wellman et al., 2001). Understanding other minds offers a rich new source of information about the beliefs and intentions that underpin others' communications, thereby helping children to fine-tune what they glean from others.

In fact, children have such a propensity to believe that informants are providing relevant information that they often repeat actions that they observe an adult perform even when those actions serve no causal purpose, a phenomenon known as over-imitation (Keupp et al., 2018). Interestingly, older children tend to over-imitate more than younger children, a tendency pointing to their growing capacity for understanding the pragmatics involved in human action, such as the idea that actions are goal-directed (Williamson & Markman, 2006; Baldwin & Baird, 2001). This is also a species-specific phenomenon, as young chimpanzees are less likely to over-imitate actions demonstrated by adult humans than young children—a finding that points to children's responsiveness to the cultural transmission of knowledge (Whiten et al., 2009). The likelihood of over-imitating is also impacted by the relationship a child has to an adult: in one study with 4-to-7-year-olds, older children were more likely than younger children to over-imitate in lab settings with a stranger, but children were equally likely across ages to over-imitate at home with a caregiver (Bach & Pauen, 2025). All in all, this evidence points to over-imitation as an adaptive propensity to learn from trusted adults, and a method children use to apply action-based pedagogy to their learning rather than relying solely on their direct causal investigations.

In addition to exploiting action-based pedagogy in their learning, children are also sensitive to the pragmatics of verbal pedagogy. In one study assessing 3-to-5-year-old children's pragmatic inferences with respect to object categories, pedagogical agents supplied an object

property (“small”) before naming an object—implying that most objects in that category were larger than the exemplar. When asked, older children were more likely than younger children to correctly identify the standard (larger) object within the same category among several options—indicating they were able to infer object categories from the pragmatic inferences involved in the verbal description of the first object (Horowitz & Frank, 2016). In another seminal line of research regarding the role of verbal pedagogy in children’s exploratory play, Bonawitz and colleagues introduced a study where an adult showed a child a novel object that afforded multiple possibilities (e.g., a tube that squeaked when pulled, a musical button, a mirror fixed inside a tube, etc.) (2011). The adult demonstrated how to elicit one such property: pulling out a tube to produce squeaking sound. What differed across experimental conditions was the lead-in that children experienced. Children were introduced to the novel object in one of four ways: with a pedagogical statement (“this is how my toy works”), a naïve statement (“Huh, did you see that?”), a baseline statement (“Wow, look at this”), and an interrupted statement, which was identical to the verbal pedagogy statement except the informant left the area directly following pedagogy. In the verbal pedagogy group, respective to those who saw the naïve or interrupted demonstration, children a) played with the target function that was introduced for a significantly greater proportion of time, b) played with the object for less time overall, c) discovered fewer main object functions, and d) performed fewer unique actions than those in the other three groups. The authors concluded that verbal pedagogy is ‘double-edged’ in that it increased children’s exploratory play with the target function that was introduced during pedagogy, but it diminished children’s broader exploration of the novel object.

In particular, there are three components indicating an adult's pedagogical competence to a child—relevant to both verbal and action-based pedagogy: their perceptual access, epistemic merit, and communication skills (Mazzarella & Pouscoulous, 2021). First, perceptual access to an object or topic one is teaching about conveys competence. For example, in one study, preschoolers were more likely to accept the claims of speakers who had perceptual access to an informative event over those who did not (Fedra & Schmidt, 2019). Findings from this study demonstrated that the ability to weigh perceptual access as an indicator of competence grows with age, as 4-year-olds were more likely than 3-year-olds to reject the claims of speakers who did not have perceptual access. The second indicator of an informant's competence is the extent to which they are knowledgeable, and the discussion of the selective trust paradigm reviewed evidence that children become increasingly sensitive to informants' knowledgeability from 3-to-6-years-old. The third component of pedagogical competence relates to an agent's ability to communicate information in a way that is useful and relevant to a child's learning. In the paragraphs above, I noted that although children are receptive to adult communication from a young age, they weigh the relevance and exhaustiveness of such communication to a greater degree as they age.

All in all, a substantial body of evidence indicates that children in preschool and beyond selectively learn from pedagogical agents and discriminate among them based upon their merit, which is conveyed by perceptual access, epistemic competence, and communicative cues. Based upon this body of evidence, one key manipulation in Study 1 of the dissertation has to do with modulating the epistemic competence of verbal pedagogy via communicative cues. As in Bonawitz et al., 2011, one agent presented a knowledgeable verbal pedagogy statement ("This is how my toy works") and one agent presented a naïve verbal pedagogy statement ("I wonder how

this toy works"). This manipulation allowed an understanding of the extent to which these dissertation findings accord with the prior literature, as well as a clarification of relations between pedagogy and the other main manipulation—the event details involved in pedagogical actions.

I now turn to discussing background evidence for the second main manipulation in Study 1, relating to the events that children observe. In the following section, I discuss how children use 'foundational' evidence in their causal learning (gained via their own sense perception, inductive reasoning, or by observing the effects of their actions), at times in combination with pedagogically-conveyed evidence and at times as a sole source of evidence supporting belief-formation. Later on, I discuss a method for examining children's event processing related to pedagogical actions called the 'dwell-time paradigm,' a central feature of Study 2 in the dissertation.

1.3 A Developmental Lens on Direct Perception and Exploration as Sources of Evidence in Children's Knowledge-Acquisition

Children's exploratory behavior is both guided by prior knowledge and aimed at gaining additional knowledge. A fundamental theory in this vein suggests that when faced with uncertainty, children both call upon prior knowledge (or "exploit" it) and actively pursue new knowledge (by "exploring") (Gopnik, 2020). The *explore-exploit* tradeoff is not exclusive to learning in childhood, but the extent to which learners exploit their prior knowledge versus explore to generate novel evidence changes with age. In one task, participants could either learn a two-dimensional rule (for example, objects with black patterns *and* objects with a certain kind of white pattern cause a light to turn on) or a one-dimensional rule (for example, objects with black patterns but not white patterns cause a light to turn on). Preschoolers were better at

learning the two-dimensional rule than adults because they were more likely to test a broader set of possible blocks as they decided which ones turned on the light (Liquin & Gopnik, 2022). This accords with what we understand about children's knowledge more generally. Since children have less prior knowledge to exploit, it is advantageous for them to be more exploratory than adults. This pattern is not exclusive to a comparison between children and adults: even across ages 3-to-6, children become less likely to explore and rely more on their prior knowledge as they age (Şen et al., 2024).

In addition to concluding that young children are especially likely to explore, a great deal of research has also focused upon the degree to which their exploration is systematic given a variety of causal evidence. In 2000, Gopnik and Sobel introduced the blicket-detector paradigm, where an object lights up or plays music only when another object (called a "blicket") is placed on it. Results indicated that preschoolers used causal information to categorize objects with a new label, even when objects otherwise differed in other meaningful features children typically use to categorize objects (e.g., shape, color). This suggests that children use the information they observe during events to recognize the causal qualities of objects even when category information is not conveyed via other means. Further research clarified that children were not simply making associations among events and the causal properties of individual objects, but they were using causal reasoning to determine the relations between the blicket(s) and the blicket detector (Sobel et al., 2004). Specifically, when children saw both A and B blocks together causing the blicket detector to light up and later learned that A on its own caused it to light up, they could discount B as a cause (even when they never saw B alone on top of the detector). Older children (age 4) were more effective at recognizing A as a standalone cause for turning on the blicket detector than younger children (age 3). Importantly for this dissertation, the blicket-

detector research identified that children use their direct observation of events—sometimes in tandem with exploratory actions—to reach causal conclusions, and their accuracy in reaching these conclusions grows with age.

In the two-and a-half decades since the original blicket-detector studies, a myriad of research has shown that children's exploratory behavior in related causal-learning scenarios is goal directed, intended to close knowledge gaps, driven by curiosity, and especially likely when evidence is ambiguous or incomplete (Anderson et al., 2023). For example, when children were trying to reach a goal (helping a character throw or jump) and they were asked to make predictions by selecting which features to see (arms or legs), children ages 3-6 years increasingly chose to reveal the most informative evidence (Török et al., 2025). This indicates that exploration is not merely a form of spontaneous information search but directed toward reaching goals by understanding the causal features involved, as seen in other experiments as well (Lapidow & Walker, 2020). Exploration is also aimed at closing knowledge gaps: children are more persistent in their searching behavior when there is a greater degree of uncertainty and more information to be gained by continuing to search (Ruggeri et al., 2023). Exploration is associated with curiosity, particularly when curiosity is sparked by surprise (Gottlieb et al., 2013). Finally, exploration seeks to clarify ambiguous evidence, and 4-to-7-year-olds explore more when their knowledge is incomplete or inconsistent (Gweon et al., 2014; Schulz & Bonawitz, 2007; Legare et al., 2010).

Challenges have been levied regarding the idea that younger children are unilaterally more exploratory than older children. In a recent study with 4- to 6-year-olds related to the double-edged-sword-of-pedagogy phenomenon described earlier, older children performed a greater number of unique actions than younger children, indicating enhanced, rather than

reduced, exploration with age (Bass & Bonawitz, 2024). This suggests the possibility that explore-exploit tendencies operate differently in terms of how age relates to the influence of pedagogy on exploratory play in tasks like that utilized in this dissertation, and this warrants further investigation in future.

Children understand and have expectations about a variety of physical properties of objects (such as solidity, continuity, their relation to gravity, etc.) (Spelke & Kinzler, 2007; Spelke, 2024). While the blicket-detector paradigm revealed how children explore when causal evidence does not clearly match with category evidence, researchers have also investigated how children test their hypotheses about causality in real time by eliciting surprise via impossible or improbable events. As mentioned above, surprise may elicit curiosity, which in turn leads to an increase in exploratory behavior. This surprise is often generated by violating children's understanding of physics and the properties of objects. I turn now to discussing this line of work and its relation to children's exploratory behavior.

1.3.1 Violation of Expectation and the Role of Surprise in Exploration

One method for understanding children's causal knowledge is to record their responses to possible and impossible events. In infancy research, the violation-of-expectation (VOE) paradigm is a longstanding method for probing infants' understanding of the physical world (for a review, see Margoni et al., 2024). Evidence from the VOE paradigm was initially deployed to investigate the extent to which infants possess basic knowledge of core physical principles, such as object permanence, at a younger age than the 18-24 months proposed by theorists such as Piaget (1954) for the consolidation of such understanding. Baillargeon, Spelke, and Wasserman (1985) showed one group of 4- to 5-month-old infants a physically plausible scenario in which a drawbridge-like screen rotated down until its movement was stopped by a box in its path.

Another group of infants viewed the same rotating screen in the context of what would be an expectation-violating scenario among those who understood object permanence: the screen rotated down and appeared to pass through the box. Infants consistently looked longer at the expectation-violating scenario rather than the plausible scenario, indicating an appreciation of object permanence. This interpretation of course hinged on the assumption that infants are especially attentive to violations of expectation, which this and a large body of other evidence has borne out (Stahl & Kibbe, 2022). Put another way, infants seem to treat expectation-violating events as a trigger to engage in extended "sampling" or information-gathering, as indicated by their sustained looking. What they glean from such sampling then presumably serves as evidence to update their prior beliefs and thus to advance knowledge acquisition.

The VOE paradigm has also been used to understand learning and exploration in preschool samples. In one study, after seeing a surprising event (during occlusion, an object seemed to have moved from under one cup to under another cup—violating spatiotemporal continuity), 3-to-6-year-olds correctly labelled a new object that contained the same properties as the one in the surprising event. However, they only correctly labelled a new object at chance levels when they had been introduced to it in the context of an expected event (an object was moved by the researcher prior to placing it under the cup) (Stahl & Feigenson, 2017). Beyond word learning, observing unexpected evidence is also associated with systematic exploration for preschool-aged children. When children saw surprising evidence (the shadow of a small object appeared larger than the shadow of a large object), they were more likely to explore in ways that allowed them to reach a relevant causal conclusion (that the distance between the objects and the light created differences in the shadow size displays) than if they saw unconfounding or irrelevant evidence (van Schijndel et al., 2015). Further, children's expectations about things

beyond spatiotemporal properties of objects can also affect their exploratory behavior. For instance, when children unexpectedly completed a difficult task that they were told most children couldn't complete, they explored more broadly than those who were told the task was easy when presented with a novel object (Doan et al., 2020).

All in all, investigations with infants and preschoolers demonstrate that they possess clear expectations about the behavior and underlying properties of objects. Violation of these expectations elicits information gathering, both in terms of extended looking as well as intensified and highly systematic manual exploration. In particular, VOE presents a unique window into children's processing of events, and findings show that viewing violating events sparks attention and curiosity. I turn now to discussing one method for understanding violations of children's expectations—the dwell-time paradigm—which is a key component of the dissertation findings in Study 2.

1.4 Examining Children's Event-Expectations Through Their Dwell Times

In the procedure for Studies 1 and 2, in addition to a verbal pedagogy introduction (as discussed above), I introduced a pedagogy-related event that in one case depicted an action consistent with children's prior expectations, and in another case violated children's prior expectations. Using an object that only afforded one function (with a large button in the center of a board and no additional features), some children saw an action likely to be consistent with their prior expectations (a finger pressed the button, and it lit up). Other children saw an action likely to violate their prior expectations (a finger waved in the air over the button, and it lit up). Importantly, we recorded children's attention to these events using a method known as the dwell time paradigm. This method involves recording the latency between mouse clicks as participants move through still images extracted from video-recorded events (Hard et al., 2011). Previous

research utilizing the dwell time paradigm documents its value -- in both children (e.g., Kosie & Baldwin, 2021) and adults (Baldwin & Kosie, 2021 for a review) -- for capturing attentional differences associated with viewers' causal inferences. In one study utilizing the dwell-time method, preschoolers viewed a canonical event (reaching over a barrier to grasp a ball) and a non-canonical event (reaching in the same action as though there was a barrier to reach a ball even when the barrier was absent) (Kosie & Baldwin, 2021). Children's dwell times surged during the reaching event in the non-canonical case but not in the canonical case, indicating that they structured their dwell times based upon what they inferred about others' goals (i.e., the latter demonstration violated the goal of reaching the ball efficiently). Utilizing the dwell-time paradigm in Study 2 of the dissertation allowed an understanding of how children responded to the violating versus consistent event, while additionally acting as an implicit measure of children's processing of the event in real time. For these reasons, the dwell-time paradigm lends strength to the study design by offering a look into children's attention to a pedagogical action directly following the verbal pedagogy introduction.

1.5 Approaching Children's Everyday Learning: Pedagogy, Event-Expectations, and Executive Function

In Study 1, children first heard the pedagogical statement and then viewed the dwell-time images for the related event (seeing either a press or a wave action). Children were then allowed to play with the object displayed in the event sequence themselves. Documenting children's unstructured exploratory play in this fashion is common within the literature related to verbal pedagogy (Bonawitz et al., 2011), violation-of-expectation (Sim & Xu, 2017) and causal reasoning paradigms more broadly (Schulz et al., 2008; Cook et al., 2011). It offers the

possibility to understand how children are testing their hypotheses in a free-form, naturalistic way that matches how learning tends to operate within their typical experience.

In their everyday learning contexts—at home, school, and in their communities—children are bringing multiple sources to bear in their learning all at once as they process a continuous stream of perception involving social agents, objects, ideas, physical laws, and all of their interactions. Still, the field has yet to investigate the effects of combining changes in two key types of evidence influencing children's learning simultaneously (verbal pedagogy and action demonstrations) while recording implicit measures of cognitive engagement involved in related processing. A primary goal of this dissertation, therefore, is to investigate the joint influence of differences in verbal pedagogy and action demonstration events on children's exploratory play while recording children's attention to the events themselves. Doing so will allow us to understand more about how children consider evidence—not as isolated or episodic, but as it appears in the real world: existing within a rich field of intermingling events that unfold across time.

A secondary aim of the dissertation is to understand how child-specific factors relate to children's processing and exploratory play within the context of these studies. In particular, children's scores on executive function (EF) and temperament measures were recorded to test their relations to the variables above. Executive function describes a core set of cognitive processes that emerge across childhood and are pivotal for children's success in learning environments, including working memory, inhibitory control, and cognitive flexibility (Blair, 2018). Previous research indicates that EF and temperament both impact how children take in new information and use it to update their beliefs in causal learning scenarios. For example, in one finding, children's ability to generate an accurate causal inference (concluding that a

hammer broke a flowerpot instead of a light bulb, paintbrush, or glue) was related to EF even when controlling for other factors (Bauer & Booth, 2019). Temperament describes children's stable emotional and behavioral tendencies. Prior work has linked temperament subscales such as persistence with positive outcomes in pedagogical contexts, and other temperament subscales such as reactivity with poorer learning outcomes in these contexts (Chong et al., 2019). In the dissertation, I mainly focus on inhibitory control in relation to EF, and the subscales effortful control and extraversion/surgency in relation to temperament.

Inhibitory control has to do with a child's ability to inhibit a prepotent response (Carlson et al., 2005). This ability could impact a variety of children's responses relevant to weighing evidence and exploring systematically, making it more likely to:

a) Suspend belief in the statements provided by an adult informant who does not seem knowledgeable (Jaswal et al., 2014).
b) Suppress the preference for choosing a salient (but non-causal) alternative explanation (Bonawitz et al., 2012).
c) Engage in goal-directed exploration rather than engaging in prepotent actions driven by salient visual object features (Gaspelin & Luck, 2019).
d) Update knowledge when prior evidence is overturned (Brod et al., 2019).

I will examine relations among inhibitory control, temperament, exploratory play, verbal pedagogy, and event-expectations in an exploratory manner given the novelty of investigating relations among these variables. In contrast, I will examine the influence of the main independent variables (verbal pedagogy and event-expectations) on exploratory play in a confirmatory manner given the body of background evidence specifically informing *a priori* predictions regarding these two variables. In particular, I employ the Head Toes Knees Shoulders task

(Ponitz et al., 2009) as an executive function measure primarily focused on inhibitory control to examine links to exploratory play, as exploratory play in the context of the current study seems to require that children inhibit a prepotent tendency to perform an expected action. Examining relations between EF and temperament to verbal pedagogy and action-expectations in this way presents another key contribution to the literature as it allows us to understand ways in which child-specific factors influence learning behaviors alongside pedagogy and prior expectations.

The dissertation focuses on children between 3 and 7 years of age. This age range has been the focus of much of the related literatures discussed above. It is a period during which children have already amassed considerable causal knowledge that they readily exploit yet engage actively in updating that knowledge through exploratory hypothesis-testing. It is also an age range during which some studies have identified an age-related shift toward reduced exploration (and thus increasing reliance on exploitation of existing knowledge), raising the possibility of novel age-related findings emerging in the dissertation findings. Lastly, the 3- to 7-year age-window was also promising from the standpoint of the method I employed in the dissertation to track children's attentional patterns during event-processing; this method has been successfully utilized in prior research from three years of age and up (Kosie & Baldwin, 2021; Meyer, et al., 2011).

1.5 Conclusion

In this chapter, I discussed children's causal learning within a constructivist framework (or child-as-scientist view), which proposes that children update their inferences by actively sampling information in their environments. The information that children sample comes from multiple sources, including direct observation of events and hearing from pedagogical agents, and children are selective in their use of evidence from these sources. Findings in this area

suggest that children's subsequent exploration in these types of learning environments is systematic and aimed toward closing gaps in knowledge and reducing ambiguity. Child-specific factors such as executive function and temperament also impact children's ability to respond to cues and explore in systematic ways that provide a foundation for belief-updating. In the following chapter, I report findings from a study involving children's causal learning within a pedagogical scenario where children either viewed a violating or non-violating event. Including both pedagogical cues and event cues allowed for an understanding of their joint impacts on children's subsequent exploratory play, and I additionally report on relations between these variables and children's executive function and temperament.

CHAPTER II

STUDY 1: THE ROLE OF VERBAL PEDAGOGY, ACTION EXPECTATIONS, AND EXECUTIVE SKILL IN JOINTLY SHAPING CHILDREN'S EXPLORATORY PLAY

Introduction

In this chapter, I report on the first of two empirical studies in the dissertation, which focuses on examining children's play behaviors in relation to the kind of verbal pedagogy they hear and the expectations they hold for an object's causal functionality. Although these distinct sources of influence on children's knowledge-updating are individually well-substantiated and widely recognized by developmental scientists, they are rarely investigated together. Yet the world ubiquitously presents itself to children as a complex combination of these differing sources of information. Gaining a sense of how children manage such complexity as they learn is thus a crucial aspect of developmental science's larger project of understanding knowledge-updating processes. With this in mind, in this study I opted to present children with learning scenarios varying simultaneously in pedagogical statements and in action demonstrations related to differences in children's prior expectations for what was likely to unfold in the events they were experiencing. Of particular interest was the way in which these two sources – verbal pedagogy and prior expectations – might jointly shape the way children processed and responded to the information presented. This investigation will take us closer to understanding how children's learning unfolds in everyday contexts, where they are typically weighing multiple types of evidence in their belief-formation.

As briefly mentioned in the general introduction, verbal pedagogy was presented to children in one of two ways in the current study: knowledgeably ("this is how my toy works") or naively ("I wonder how this toy works"). I chose this verbal pedagogy manipulation as it is in line with known effects from the previous literature (for example, the *knowledgeable verbal*

pedagogy statement is identical to that used by Bonawitz et al., 2011, where it was related to constrained exploration). The other form of verbal pedagogy I employed – *naïve verbal pedagogy* – contrasted in terms of the adult's apparent knowledgeability concerning the workings of the object they seemed to be teaching about. As I discussed in the previous chapter, the knowledgeability versus naivete contrast tends to have strong effects on children's judgements about the reliability of pedagogical agents.

Children in the study then experienced one of two events following the introductory verbal pedagogy statement; these events differed in relation to children's likely expectations for the causal functionality of the object involved. I chose to design an object that visually afforded only one function (lighting up), but contained multiple possible causes for that function. In other words, the object could be acted upon in multiple ways to produce the same outcome. In one demonstration – the *consistent expected action* condition --pressing the button caused the light to turn on, an action that would likely be consistent with children's prior expectations as they probably had seen many buttons turn on many lights before. In other words, a pressing action to turn on the light should be a probable cause-and-effect sequence. In the other demonstration – the *violating expected action* condition -- waving over the button caused the light to turn on, an action that ought to violate children's prior expectation that pressing would turn on the light (the idea being that waving is a possible, but improbable, cause for turning on lights in children's prior experience). As discussed in Chapter 1, a great deal of work has assessed how children respond to violating stimuli. This review suggested that observing a violating event may prime children to pay greater attention to details of the learning situation that follow—and this effect may be especially pronounced when children's "core" expectations about the world are violated (Stahl & Feigenson, 2019). Similarly in this study, I set out to understand how presenting an

action that violated children's prior expectations about an object (waving) would influence their likelihood of exploring the object to a greater degree than those who saw an action that was consistent with their prior expectations (pressing).

As in prior research, I expected that hearing *naïve verbal pedagogy* (as opposed to *knowledgeable verbal pedagogy*) would be associated with increases in children's exploratory behavior—namely, the number of unique actions that they would perform (one of the primary findings from Bonawitz et al., 2011). However, I expected that the contrasting forms of verbal pedagogy (knowledgeable versus naïve) would relate to the proportion of time children spent enacting the demonstrated action—one of the other play outcomes from the 'double edged sword' study—somewhat differently than was observed in prior double-edged sword research. Specifically, when children saw an action consistent with their prior expectations (i.e., pressing to turn on the light), I expected the typical double-edged sword finding for this measure (i.e., enhanced proportion of time enacting the demonstrated action in response to knowledgeable relative to the naïve verbal pedagogy). However, in the case when children saw a violating action, the proportion of time children spend with the demonstrated action should reflect not only the influence of verbal pedagogy, but also children's ability to flexibly update their understanding of the object's causal functionality. Therefore, I predicted that while hearing knowledgeable verbal pedagogy might still orient children's focus to the demonstrated action, children's executive function might also influence the extent to which children devoted time to performing the demonstrated action.

Related to the impact of the expected action on children's exploratory play, I expected that across groups, children would perform the action that was consistent with what the object afforded (and therefore, more salient than the violating action). I expected that children would,

by and large, only perform the violating action when they saw it performed by a pedagogical agent. Finally, I expected verbal pedagogy and expected action factors to simultaneously interact, such that those who heard the naïve verbal pedagogy and saw the violating action would show an especially high degree of exploration in their play.

In what follows, I investigate how both verbal pedagogy and the probability of an action demonstration together influenced children's exploratory play by addressing the following research questions:

Confirmatory

1. To what extent will children's responses to a knowledgeable versus naïve verbal pedagogical introduction shape children's exploratory actions? To what degree will these exploratory play patterns replicate and extend the 'double edged sword of pedagogy' findings?
2. To what extent will action expectations shape the exploratory actions children display?
3. To what extent will action expectations and verbal pedagogy interact in their impact on children's exploratory play?

Exploratory

4. To what degree will children's executive function (EF) skills display significant associations with their patterns of exploratory play when controlling for age?
5. To what degree will children's EF skills interact with verbal pedagogy and action expectations in their association with exploratory play patterns when controlling for age?
6. To what extent will parental responses to the CBQ scales Effortful Control and Extraversion/Surgency show significant relationships to children's exploratory play behaviors?

7. To what degree will CBQ scales Effortful Control and Extraversion/Surgency interact with verbal pedagogy and action expectations in their relationship to children's exploratory play?

Method

2.1 Participants

The full sample included 223 participants recruited from the floor of the Oregon Museum of Science and Industry (OMSI) in Portland, OR and the Eugene Science Center (ESC) in Eugene, OR. Based upon similar research in this area (Kosie & Baldwin, 2021), I performed an *a priori* power analysis using the "pwr" package in R (Champely, 2023) with the goal of obtaining 0.9 power to detect moderate effect sizes (0.25 and above). This test revealed an ideal sample of roughly 58 participants per condition (N = 231). The final sample (55 participants per condition) attained a similar degree of power (0.88) while reflecting minor constraints related to museum scheduling and research assistant availability post June 2025. Three participants over 7.5 years (90 months) were initially excluded, and the final sample included 220 participants (*M*age = 59 months; range = 2.8-7.3 years, 111 female). Race was representative of the area (69% White, 11% Asian, 11% Biracial, 2% Black, < 1% Indigenous American). Most parents (74%) reported holding a Bachelor's degree or higher (see Table 2.1). Demographic data, parent education, and subjective social status (see details in Appendix A, and Table 2.2) were missing for eleven participants whose parents did not fill out the form.

Table 2.1. *Parent education level as reported by the parent filling out the survey.*

Parent Education Level	N (% of sample)
High School or GED	9 (4.1%)
Some college (no degree)	30 (13.6%)
Associate's degree	14 (6.4%)
Bachelor's degree	69 (31.4%)
Master's degree	58 (26.4%)
Doctorate degree	10 (4.5%)
Professional degree	19 (8.6%)
No response	11 (5%)

Table 2.2. *Parent-reported Subjective Social Standing from the MacArthur Scale of Subjective Social Status (Adler et al., 2000). Parents placed themselves on a social standing ladder ranging from 1-10 for the U.S. and for their Community. See Appendix A for details.*

Score	USA: N (% of sample)	Community: N (% of sample)
1	0 (0%)	0 (0%)
2	0 (0%)	1 (0.5%)
3	7 (3.4%)	7 (3.3%)
4	17 (8.2%)	8 (3.8%)
5	37 (17.8%)	41 (19.6%)
6	31 (14.9%)	43 (20.6%)
7	60 (28.8%)	47 (22.5%)
8	37 (17.8%)	43 (20.6%)
9	14 (6.7%)	14 (6.7%)
10	5 (2.4%)	5 (2.4%)

2.2 Materials

2.2.1 Novel Object

An object was created for the study which was meant to be relatively novel for children. It had a black base measuring 5.5 inches x 7.5 inches x 1 inch with a 3-inch diameter white light encasement set in the center that was roughly 1.5 inches tall, with a central region -- approximately 2 inches in diameter -- that lit up. When turned on, the light was red. The light was non-operable by the user except via remote control. Research assistants stood behind and to the side of the table and remotely activated the light every time the child either pressed the top of the central region or waved over it. No children noticed or made comments about the light being turned on remotely by one of the researchers or by trickery. The novel object can be viewed in the videos and slideshows available in the supplementary materials: https://osf.io/h6wat/files/osfstorage

The novel object met the criteria for this study in terms of possessing the capacity to be acted upon to produce a causal outcome which can be achieved by more than one action. Research assistants remotely activated the light every time a child attempted a press or a wave, regardless of condition. The light remained on until about 0.5 seconds after a press or a wave action had ended. In cases where children rapidly and continuously pressed or waved, the light remained on while periodically flickering. The light could be on while the object was in any orientation (e.g. sitting on the table, upside down, held in the air, partially picked up, sideways); however, if a child moved the object below the tabletop, they were directed to move it higher so it would be visible in the recorded video. Less than ten children ever needed this reminder.

There is precedent from prior developmental research both for using an object that is controlled by a researcher and for using an object with a single function similar to ours. For

example, in the blicket detector paradigm (Gopnik & Sobel, 2000), researchers turned switches on and off underneath a table hidden from the child's view to give the impression that particular objects caused an effect while others did not. Another series of studies (Meltzoff, 1988; Gergely et al., 2002) utilized an object very similar to ours in order to investigate infants' understanding of the pragmatics of action via their likelihood of over-imitating. Although the original 'double edged sword' object (Bonawitz et al., 2011) contained four key functions, the fact that our novel object included just one function (light) allowed us to investigate the role of action expectations within a more interpretable scope given multiple independent variables.

2.2.2 Verbal Pedagogy Videos

Two videos were created to convey verbal pedagogy about the function of the novel object. In each video, an adult was head-on to the camera with arms by their side while sitting in front of a table. The novel object that was described in the section above sat in front of them on the table. In one video, the adult made a knowledgeable statement, "This is how my toy works." In the other video, the adult made a naïve statement, "I wonder how this toy works." Each video was roughly 5 seconds long. Efforts were taken to keep the appearance of the person and their prosody consistent across the two videos such that they varied only in verbal pedagogy content with slight expressional differences to match the message of the words (for example, raised eyebrows and slightly tilted head in the naïve case). These videos are available for viewing in the supplementary materials: https://osf.io/h6wat/files/osfstorage

2.2.3 Expected Action Slideshows

Two slideshows were created through extracting slides from corresponding digital videos. The videos depicted an actor performing an action on the novel object described above. Both videos were recorded at a rate of 30 frames per second. In both videos, the torso and hands

of the actor are depicted above a tabletop (the actor was the same individual and the general setting was the same as that depicted the verbal pedagogy videos described above). The object is held aloft at the start of both videos. In the *"consistent" expected action* video, the right hand of the actor leaves the object from the left-hand side of the screen and makes an arcing motion before pressing—and turning on—the light. The right hand then returns to grasping the right side of the object; the left hand holds the base throughout and never moves. The *"violating" expected action* video depicts a motion where the right hand arcs over to a little beyond the left side of the object, then changes trajectory to perform a horizontal waving motion visibly above the light encasement. The light turns on despite the finger obviously never touching it. The right hand then returns to grasping the right side of the object; the left hand holds the base throughout and never moves.

I made every effort to equate, as much as possible, the precise movements depicted in the consistent and violating action videos (and hence also the slideshows). For example, the actor moves at generally the same pace (resulting in videos of roughly the same length), activates the light for generally the same duration, and performs all minor, supportive actions in an identical fashion (e.g., holding the base with two hands to start and maintaining hold of the base with the left hand throughout the press or wave action, thereafter grasping the base with the right hand)., Importantly, the actions did not differ from one another in the degree to which they were canonical examples of their action type; specifically, both press and the wave actions were non-canonical in the sense that the arc of the arm shown during the press and the arc of the arm shown during the wave were both an exaggerated form of action.

Each video was recorded at 30 frames per second, and was roughly 4 seconds long. Still images were extracted at regular intervals of 5 frames for each video, resulting in 21 total images

per event. Both slideshows depicted the light turning on at slide 10 and turning off at slide 14 (see Figure 2.1). Automated versions of the slideshows of both events can be found in the supplementary materials: https://osf.io/h6wat/files/osfstorage (note that in the study itself, children advanced through the slideshows at their own pace). The computer portion of the task that included these slideshows took place on a Mac laptop using PsychoPy® software (Pierce et al., 2019). Participants wore child-sized noise cancelling headphones and clicked a child friendly mouse throughout the computer task. In a handful of cases when headphones were bothering participants, they were taken off following the verbal pedagogy video, which contained the only audio element in the task.

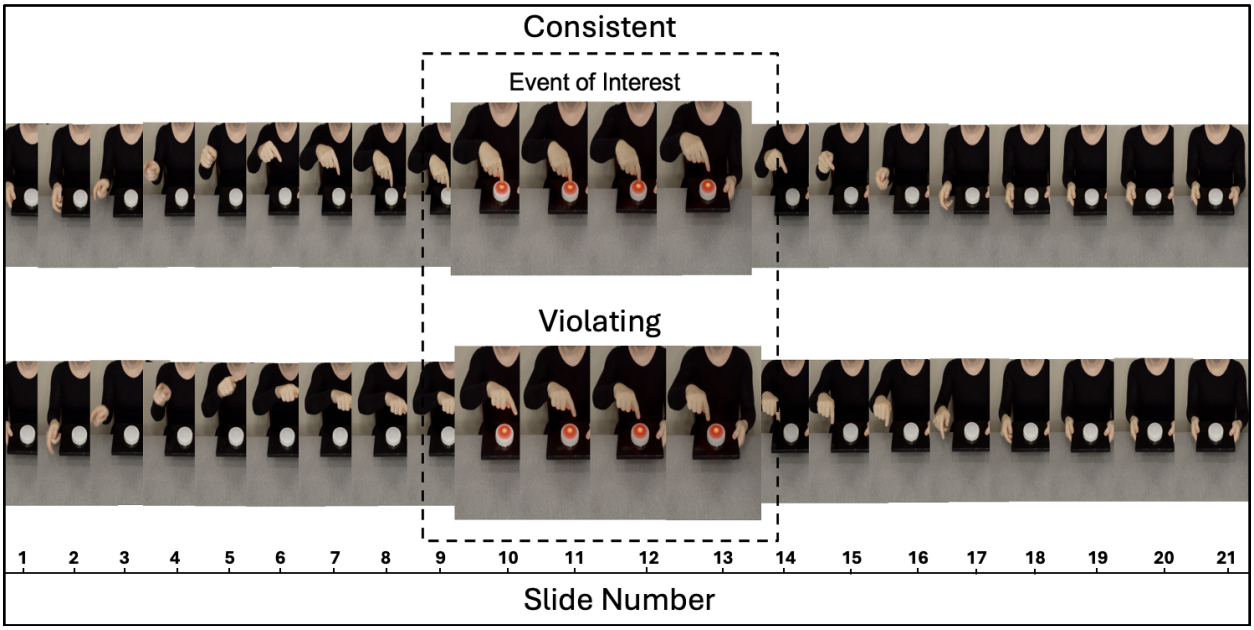


Figure 2.1. Expected action slideshow images for consistent (press) and violating (wave) events. For each slideshow, the light is on from slides 10-13.

2.3 Design

Participants were recruited in accordance with IRB ethical guidelines as well as National Living Lab guidelines; The Living Lab is a National Science Foundation initiative for the promotion of developmental science research and science communication in museum settings (National Living Laboratory® Initiative). This means that every child who expressed an interest

in the task was allowed to participate. If a parent or guardian was not available to consent to their participation in the research, no data were collected (I note that this was only the case for a handful of participants). Additionally, researchers were trained to provide a debrief and to engage in dialogue with parents and children about the research questions involved in adherence to the National Living Lab model.

Participants were randomly assigned to one of 4 equally sized, pre-determined conditions, counterbalanced by age and gender (randomization took place in base R prior to recruitment using the sample function with replacement). As described earlier, in the *knowledgeable verbal pedagogy* condition, children saw a video with a knowledgeable informant who made the statement: "This is how my toy works." In the *naïve verbal pedagogy* condition, children saw a similar video with a naïve informant who made the statement "I wonder how this toy works." Following the verbal pedagogy video, participants advanced at their own pace through either the *consistent expected action* slideshow (in which the depicted action – a press that turned on the light – was consistent with their prior expectations for action on the novel object) or the *violating expected action* slideshow (in which the depicted action – a mid-air wave that turned on the light – violated their expectations for action on the novel object). Both verbal pedagogy and expected action slideshow were therefore between-subject independent variables, resulting in a 2 X 2 factorial design (See Figure 2.2 for details). Slideshow viewing (First vs. Second) was the only repeated measure (as children viewed identical slideshows twice); however, it is not included in the analyses for this study, but in Chapter 3.

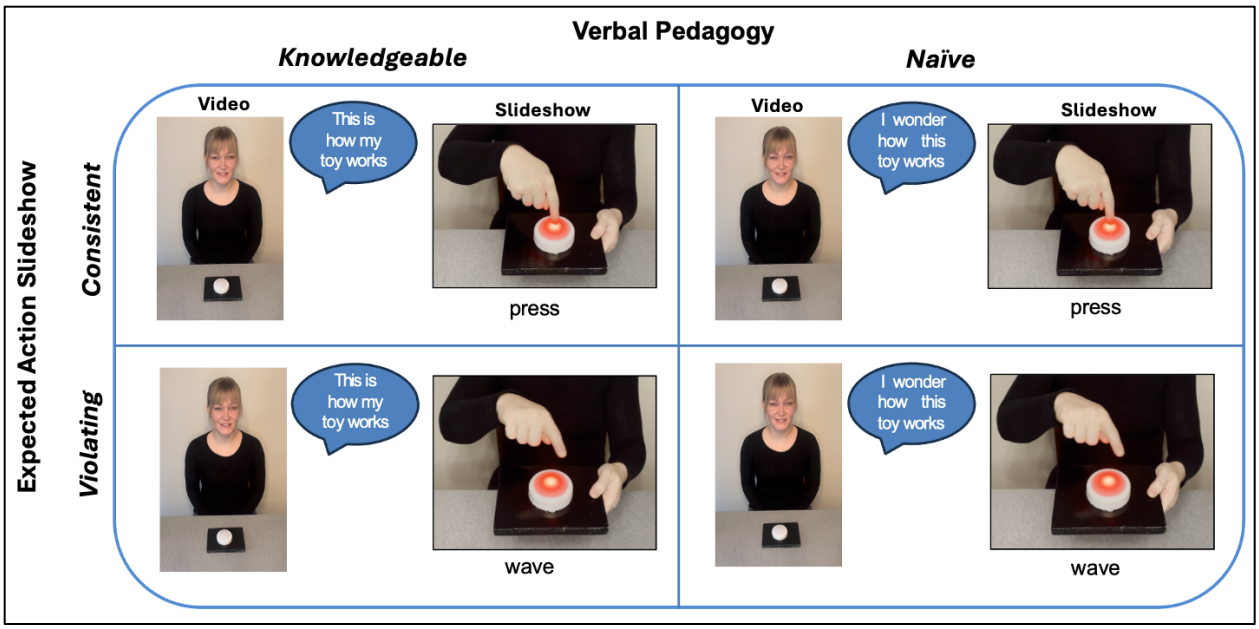


Figure 2.2. The design for the current study, including levels for the independent variables Verbal pedagogy' and 'Expected Action Slideshow.'

Measures

2.2.4 The Children's Behavior Questionnaire

The Children's Behavior Questionnaire is a measure of temperament assessing 3-8-year-old children's self-regulation skills (Rothbart, 1996). Parent responses were collected on the Children's Behavior Questionnaire Short Form (CBQ-SF) (Putnam & Rothbart, 2006) for the following sub-scales: Approach, Activity Level, Attentional Focusing, Impulsivity, Inhibitory Control, and Shyness. For all questions, parents rated their child on a 1-7 Likert scale with 1 being "extremely untrue" and 7 being "extremely true" across 6-7 items related to each sub-scale (37 questions total). For more details, see Appendix B.

2.2.5 Head Toes Knees Shoulders Task

The Head Toes Knees Shoulders (HTKS) Executive Function (EF) task (Ponitz et al., 2009) is designed to evaluate 3-7-year-old children's ability to inhibit a prepotent motor response. I was particularly interested in how children's performance on this task would relate to their exploratory play behaviors. In the HTKS task, children were instructed to respond with the

opposite movement when given a command (i.e., when asked to touch their head, the correct response was to touch their toes, and vice versa). After 10 trials with head and toes commands, participants were introduced to an additional two commands ("touch your knees," "touch your shoulders"), and there were 10 trials with all four commands. Each correct response was worth 2 points. If children initially made a movement to the incorrect location and self-corrected, they earned 1 point, and movements to incorrect locations earned 0 points. There were a total of 6 practice items and 20 test trials, and scores ranged from 0-52. According to the HTKS task guidelines, if children were in kindergarten or older, they were led through both parts of the HTKS task. If children were under kindergarten age, they started with part 1, and if they responded correctly on at least half of the test items, then they went on to complete part 2. See Appendix C for more information.

2.4 Procedure

The parents of participants were initially approached by a researcher who briefly explained the study and asked if they and their child wanted to participate. After parents had filled out a consent form and the child provided verbal assent, the parent was given an iPad to fill out the demographic and CBQ-SF surveys and asked to remain seated nearby. Children were directed to a table situated in the corner of the museum (or, at ESC, in a separate classroom). A room divider was set up blocking the front and side of the table, and the other side of the table sat along a wall; thus, a view of the table, computer task, and novel object were generally occluded from the view of others. As well, the divider helped to reduce distraction when children were participating in the study. Only one side of the table remained open (for entry), and there was one child-sized chair for the participant and one adult chair occupied by the researcher. A tripod and video camera were set up to record a front-view of the task, and another camera was

situated to record a side-view of the task. Having both camera views was particularly important to provide information about whether, and when, children might look away during slideshow viewing, as dwell times related to such look-aways were recorded to determine possible data exclusion—I discuss this more in Chapter 3.

The task began with a training period on a computer where participants learned to click through images to view events sequentially (for details, see 'Materials' in Chapter 3). Following this training, participants viewed a short verbal pedagogy video, where a novel object was introduced in either the *knowledgeable* or *naive* condition, as described above. Next, children clicked through the expected action slideshow images, which depicted the same novel object lighting up either when pressed (*consistent* condition) or when waved over (*violating* condition). To complete the computer task, children viewed the same slideshow a second time (*second viewing*).

Children advanced through slideshow images at their own pace by clicking a computer mouse with two large buttons that was designed for young children to ease their use. When a slideshow was complete, the researcher guiding children's use of the computer (Diana Daschel) advanced to the next portion of the task via the space bar at set intervals (post-training slides, before and after the verbal pedagogy video, before and after the dwell time slides). While participants watched the verbal pedagogy video and clicked through the slideshow images, the researcher turned away and wore dark glasses to occlude their vision in the service of keeping the researcher blind to the experimental condition in which children were participating.

Once they finished viewing the slideshow images, children were then offered the same novel object that they had seen displayed during slideshow viewing and instructed to let the researcher know when they were finished playing with it. Following similar guidelines as

Bonawitz et al. (2011), the play period ended in one of three ways: a) participants either self-elected to finish the play period as originally instructed, b) three minutes elapsed and the play period was ended by the researcher, or c) a child stopped playing with the toy for 5 seconds, leading them to be asked "are you finished?" and the play period ended if they responded affirmatively. If they responded negatively, they were allowed to continue playing either until 3-minutes was up, they elected to finish, or they stopped playing with the toy for 5 seconds an additional time.

Following the play period, children were asked two follow-up questions indexing their memory for what they experienced during slideshow viewing: 1. "Remember before you played with the toy, you saw a person with the toy: did the light turn on when the person pressed right on the toy, or did it turn on when the person waved in the air above the toy, or both?" and 2. "Remember when you were playing with the toy: did the light turn on when you pressed right on the toy, or did it turn on when you waved in the air above the toy, or both?" These questions were counterbalanced both in the order that a) the subject of the question (researcher vs. participant) was first presented, and b) the action (light or wave) mentioned was first presented. Children were then directed to a nearby area where a research assistant led them through the Head-Toes-Knees-Shoulders task. The decision to include this task at the end of the procedure (as opposed to counterbalancing, with some children experiencing it earlier and others experiencing it later in the procedure) was made to avoid possible effects of HTKS on children's behavior for other elements of the procedure. In particular, since the HTKS task has a pedagogical element (i.e., the researcher instructs children on what to do when), it seemed possible that participation in HTKS first might have primed children to assume that the rest of their experience in the computer task would be pedagogical, thereby undercutting the power of

the manipulation (i.e., *knowledgeable* vs. *naïve* verbal pedagogy introduction conditions). To finish, children selected a sticker as their reward and researchers offered a debrief to both parents and children (See Figure 2.3 for the procedural flow of events).

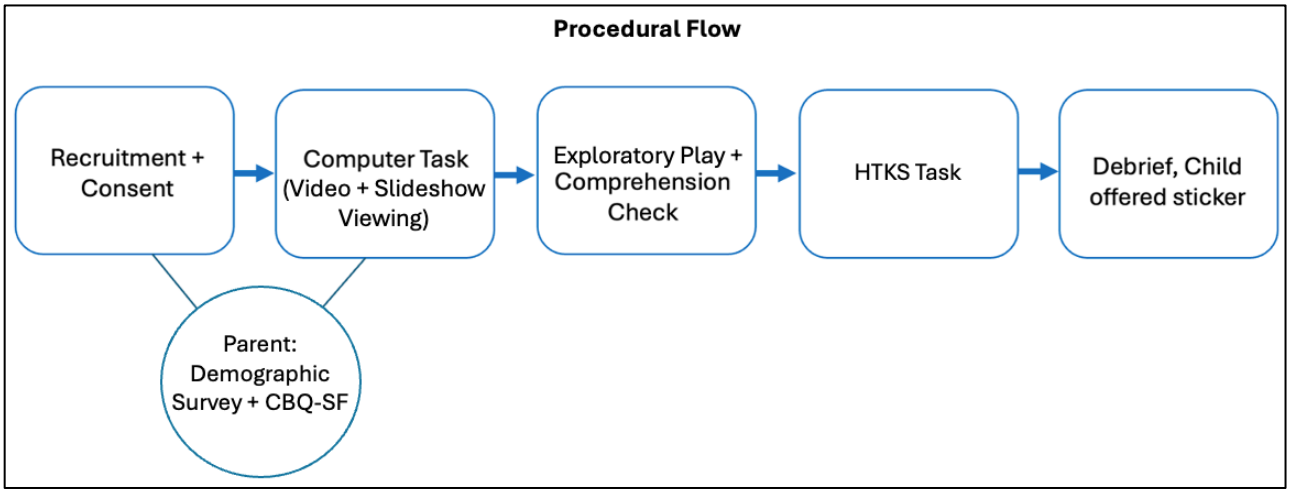


Figure 2.3. The full procedure for the current study.

2.5 Coding and Dependent Measures

In partnership with a team of undergraduate coders and advising researchers, I designed a behavioral coding systems to capture measures from the exploratory play session, memory check, and the HTKS task, as described below.

2.5.1 Play Behaviors

Video coding for play behaviors focused on the following outcomes: total time spent playing (in seconds), whether any waves were attempted (Y/N); whether any presses were attempted (Y/N); the proportion of time spent waving throughout the whole play period (waving time/total play time); the proportion of time spent pressing throughout the whole play period (pressing time/total play time); the number of unique actions children displayed during exploratory play (such as looking underneath the toy or holding it upside down) recorded as type rather than token; total actions performed (any wave + any press + number of unique actions); and the first action performed (press, wave, or other). This coding system was designed based

upon the original ‘double edged sword’ study (Bonawitz et al.), and I corresponded with the author to understand their original coding system. The ‘double edged sword’ coding system entailed most of the outcomes listed, but our team also coded the first action, any presses, and any waves. In an important departure from Bonawitz et al., while the proportion of time with demonstrated action measure is indicative of constrained exploration in the case of the consistent expected action group (as it was for them), it is indicative of children’s ability to flexibly update in the case of the violating expected action group. This is because children who had their expectations violated had to update their notions of what the object afforded in addition to adhering to the verbal pedagogy. Each exploratory play session was coded by a minimum of two research assistants (one standard coder and at least one additional coder) who were trained on the system and met regularly with Diana Daschel, who also acted as an additional coder. Coders were blind to condition while coding. When assessing inter-rater reliability, Krippendorff’s alpha reached 0.90 or above for all coding measures.

2.5.2 Memory Check

From the comprehension check questions, children responded to a question about the researcher “… did the light turn on when the person pressed right on the toy, or did it turn on when the person waved in the air above the toy or both?” as a memory check. Participants were given a memory score of ‘1’ if they were in the *consistent expected action* condition and reported that the person pressed and if they were in the *violating expected action* condition and reported that the person waved. If the participant inaccurately reported seeing a press or a wave, or if they reported seeing both, they were given a memory score of ‘0’.

2.5.3 HTKS

As instructed in the materials for the HTKS task, researchers who administered the task also scored the participant at the same time. In an effort to retain the highest degree of accuracy given that HTKS was the main EF outcome, we also video recorded these sessions and set up a *post hoc* coding system to record children's responses. I opted to use scores based on this video-based coding rather than scores from the researcher recordings during task administration. Although scores generally remained within a point or two of the original scores following video coding, a substantial portion of the videos (46%) had a change in overall score. Two expert coders recorded participant's behaviors according to the original instructions of the HTKS guidelines, and participants received either a '0', '1', or '2' for each training and test item. The Krippendorff's alpha value assessing the inter-rater reliability between these two coders was 0.96.

The full set of HTKS scores were missing data for 24 participants who did not make it through the initial training portion of the task (all but one of these participants was under the age of 4). Accordingly, the full sample for the HTKS data reflects N = 199 participants. I initially considered creating a composite EF score made up of participants' HTKS scores and their Effortful Control scores from the CBQ. I chose not to do this (but rather, to use HTKS as the main EF variable and treat Effortful Control as a separate variable) for a few reasons. First, the two were only weakly correlated, $r(179) = .25$, $p < .001$ (note that the degrees of freedom value here also reflects missingness from the CBQ data and age-related exclusions noted in the results section). Secondly, there is evidence that behavioral EF tasks and parent-report EF measures may assess different constructs as they often bear a weak relation to one another and map on to

different developmental processes (Eycke & Dewey, 2016; Silver, 2014; Nagar & Gupta, 2017; Vriezen & Pigott, 2002; Dekker et al., 2017).

Results

The analysis plan was preregistered at https://osf.io/bt235/overview. While the pre-registered analysis plan remained consistent with the reporting in the dissertation, a number of small updates were made to the approach that diverged from the pre-registered plan for other criteria. Those changes are recorded in Appendix D. All analyses were performed using R statistical software (Version 4.5.1; R Core Team. 2023). Tables were created using kableExtra (Zhu, 2024) and figures using ggplot2 (Wickham, 2016). Linear mixed effects models were estimated with the lmerTest package in R (Kuznetsova, 2017), using Sattertwaite's method to approximate degrees of freedom. The logistic regression function 'glm', general linear function 'lm', and anova function 'aov' were used from the stats package—a component of base R. Zero-one beta regression analyses were estimated with the glmmTMB package (Brooks et al., 2017). Pairwise comparisons were conducted using either the emmeans package (Lenth, 2024) to approximate estimated marginal means with a Bonferroni correction, or Tukey's Honest Significant Difference (HSD) test. Alpha was set at a standard value of .05 to detect statistical significance.

Three participants were excluded prior to analyses because their parents did not consent to recording video data (data for play outcomes required coding videos of play behaviors). Going forward in this section, the full sample contained N = 217. For age-related analyses, only 214 participants were included since the sample contained only those over three and under seven years (36-83 months), eliminating three participants (who had not already been excluded by the initial age cutoff of 90 months). I chose to do this because such a small number of participants

would fail to provide meaningful estimates for the discrete year age groups they comprised (2 years and 7 years). Although age was treated as a continuous variable in the included analyses, I sometimes report age-related data—for example, in tables and figures—by younger (3-4-year-olds) and older (5-6-year-olds) age groups or by discrete year groupings.

2.6.1 Specifying Exploratory Play Outcomes and Executive Function Measure

The Bonawitz et al. (2011) original study contained four main outcomes: 1) total play time, 2) number of unique actions participants performed (as type not token), 3) the proportion of time participants spent only with the demonstrated function, and 4) total number of target actions discovered (4 possible). Additional related studies (Zsoldos & Kiraly, 2025; Schneidman et al. 2016; Yu, Landrum, et al. 2018; Brezack et al., 2021) indicated two main outcomes were involved in a consistent replication of the pattern of findings (the number of unique actions and proportion of time spent with the demonstrated function). Accordingly, in the confirmatory analyses, I chose to focus on these two play outcomes and their relations to our independent variables. The number of unique actions is understood to represent the degree to which children's play demonstrated enhanced exploration, while the proportion of time spent with the demonstrated function represents the degree to which children's play indicated constrained exploration. Because of this, I expected that unique actions and proportion of demonstrated action outcomes would be negatively correlated. In fact, there was a pre-registered plan to perform a MANOVA given a predicted moderate negative correlation between these two outcomes. The two outcomes were negatively correlated, $r(215) = -.26$, $p < .001$, though not moderately. Thus, I went ahead with a plan to run separate analyses for each of the two play outcomes where relevant. For all analyses reported below, when I reference executive function

(EF), this corresponds to children's scores on the HTKS task. For research questions 6-7, I use the Effortful Control variable from the CBQ; however, I never label this variable as EF.

2.7 Preliminary Analyses

2.7.1 Descriptive Play Findings: The Salience of Pressing

Presses were attempted in 100% of cases for those who saw the action that was consistent with expectations (press). Presses were also attempted between 81-83% of cases for those who saw the action that violated expectations (wave), depending on verbal pedagogy group. This indicates that regardless of whether participants saw an action that was consistent with or violated expectations, as predicted the press remained a salient expected causal action that was visually afforded by the object. Waves, on the other hand, were attempted in 50% of cases for those who saw the *violating expected action* slideshow. Waves were attempted in roughly 9% of cases for participants who saw a *consistent expected action*. On the face of it, this may seem like a sizable percentage for an action that a) violated expectations, and b) was not demonstrated for these participants. This finding may be related to the fact that research assistants remotely activated the light in all cases when participants performed a waving action over the object. In other words, this group of participants may have inadvertently discovered that the light was activated by waving simply because they observed that the light turned on even if they happened to wave over the object without intending to. Even for those who saw a wave in the *violating expected action* condition, over half (57%) of children still attempted a press first, with 35% attempting a wave first, and 8% attempting another action first (such as picking the toy up and flipping it over). See Table 2.3 for descriptives related to pressing and waving behavior by age and condition. According with our pre-registered hypotheses, pressing occurred across expected action groups, whereas waving mainly occurred in *the violating expected action* group.

Table 2.3 *Proportion of participants who pressed or waved at least once by condition and age group.*

Condition	Age (years)	N	Proportion Pressed	Proportion Waved
Knowledgeable x Consistent	3–4	30	1.00	0.03
Knowledgeable x Consistent	5–6	23	1.00	0.13
Knowledgeable x Violating	3–4	29	0.97	0.21
Knowledgeable x Violating	5–6	24	0.62	0.75
Naive x Consistent	3–4	32	1.00	0.00
Naive x Consistent	5–6	25	1.00	0.20
Naive x Violating	3–4	29	0.86	0.38
Naive x Violating	5–6	22	0.77	0.77

To assess the extent to which children pressed for a significantly greater proportion of time on average than they waved, even when they never saw a press demonstrated, a linear mixed effects model was fit predicting the proportion of time spent playing based on fixed effects for action attempted (press vs. wave) and condition, plus an interaction between them, and a random intercept for subjects. The omnibus results revealed a main effect of action attempted (press vs. wave), $F(1, 426) = 436.47$, $\eta p^2 = .51$, 90% CI [0.45 , 0.55], $p < .001$ (i.e., presses were attempted for a greater proportion of time ($M = 0.69$, $SD = 0.33$) than waves ($M = 0.13$, $SD = 0.27$)), whereas no significant main effect of condition emerged. However, the interaction between condition and action was statistically significant, $F(3, 426) = 29.69$, $\eta p^2 = .17$, 90% CI [0.12 , 0.22], $p < .001$. A simple-effects analysis revealed that this interaction was driven by the fact that the difference between proportional press and wave behavior was significantly larger for the consistent expected action groups (*knowledgeable x consistent M*-diff = 0.79; *naïve x consistent M*-diff = 0.82) and smaller for the violating action groups (*knowledgeable x violating M*-diff = 0.35; *naïve x violating M*-diff = 0.28), $\Delta = 0.497$, $SE = .05$, $t(213) = 9.39$, $p < .001$ (see Figure 2.4).

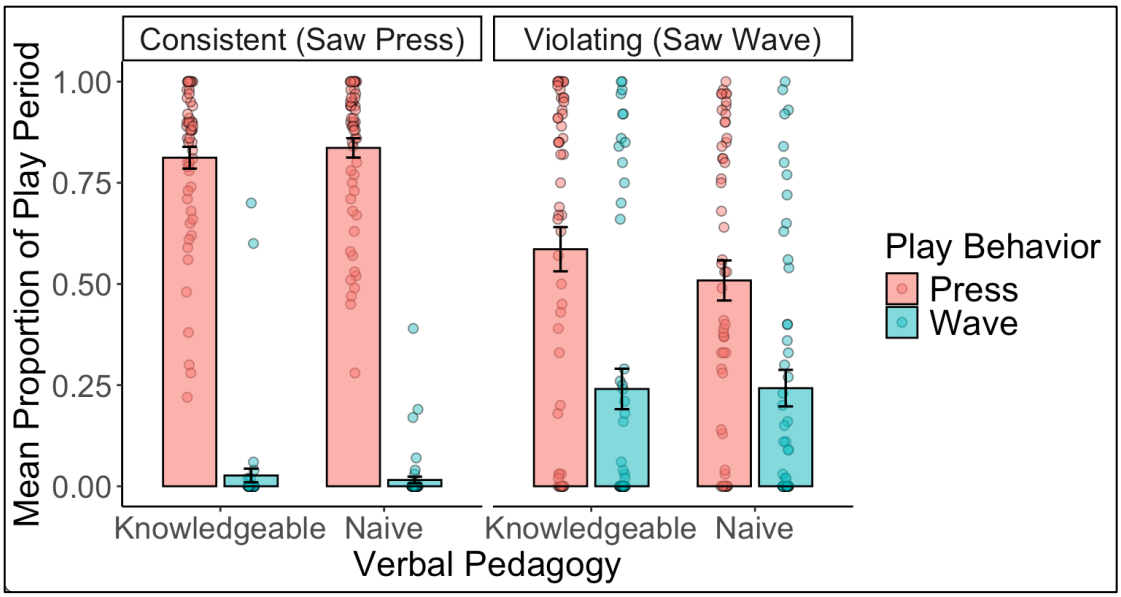


Figure 2.4. Mean proportion of the play period that children spent pressing versus waving by condition, with error bars.

2.7.2 Impacts of Age, Memory Check, and Gender on Exploratory Play

When included in analyses with the main independent variables (verbal pedagogy, expected action) for the two main play outcomes (proportion of time with demonstrated action, number of unique actions), gender produced neither main effects nor interaction effects and thus I did not include it in further analyses.

As expected, age and EF were significantly and strongly positively correlated, $r(191) = .72$, $p < .001$. Given that I was especially interested in the relationship between individual differences in EF and exploratory play outcomes, I opted to focus on main effects of EF while controlling for age in general for the reported results. In one case (in the reporting related to research question 4), I investigated a significant interaction between age and EF scores. In cases where I investigated impacts of EF while controlling for age (related to research questions 4 and 5), multicollinearity was a potential concern within the models given the strong positive correlation between these two variables reported above. However, concern about multicollinearity was mitigated because, in all cases, the Variance Inflation Factor (VIF) – a

measure of multicollinearity – was less than 2.25, indicating a degree of inflation for standard error values but the absence of severe multicollinearity.

As described above, I also recorded children's memory accuracy in a follow-up question after they viewed the slideshow, asking them whether they saw a press, wave, or both. Age was significantly related to children's memory accuracy, $r(211) = .21$, $p < .01$, though only weakly. However, as seen in Figure 2.5 and Table 2.4, participants had some difficulty in accurately remembering the action demonstrated during the slideshow, especially younger children in the *violating expected action* group (only 28% of whom accurately remembered the action they had seen in the slideshow). Because parsing the data to investigate how memory accuracy influenced play outcomes decreased sample sizes substantially, I opted not to utilize memory accuracy as a variable in the results for this study.

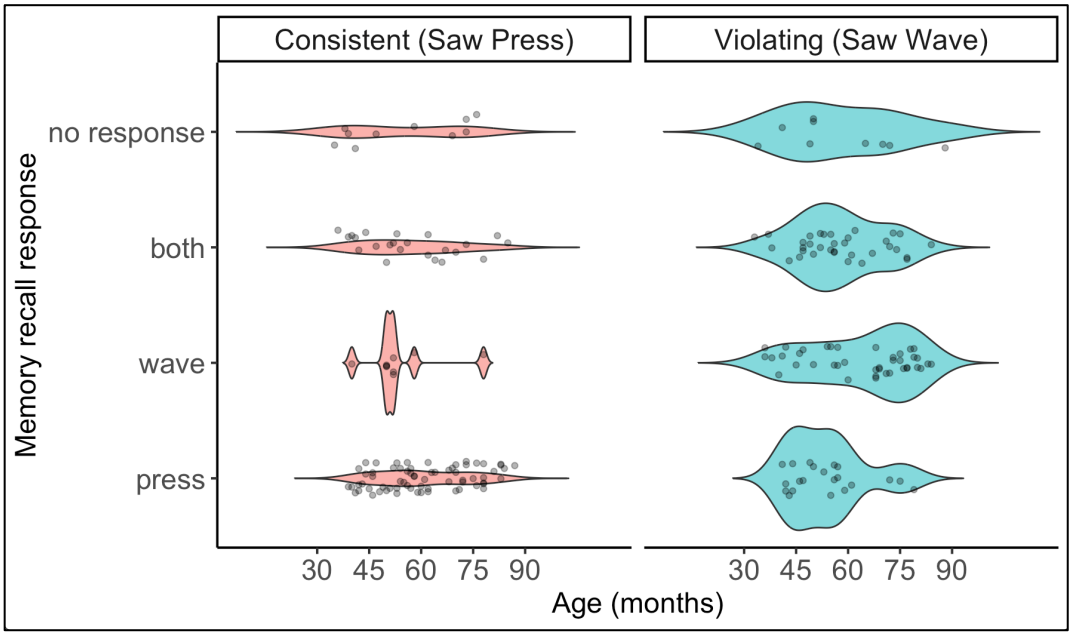


Figure 2.5. Responses on the memory recall question by age and expected action group.

Table 2.4. *The number and proportion of participants who remembered the slideshow action accurately by age and expected action group.*

Age (years)	Expected Action	N	Memory Accurate	Memory Not Accurate	Proportion Accurate
3–4	Consistent	62	34	27	0.55
3–4	Violating	58	16	41	0.28
5–6	Consistent	48	32	15	0.67
5–6	Violating	46	24	21	0.52

2.7.3 The Relationship between Age and Number of Unique Actions

As mentioned in Chapter 1, I wondered about the extent to which outcomes that index exploration (in our case, unique actions) show relations to development—in keeping with the explore-exploit phenomenon. In order to assess the possibility of age differences in children's performance of unique actions (those beyond presses and waves) a linear regression was fit containing age as a fixed effect. Results indicated no association between age and the number of unique actions that children performed, $\beta = 0.001$, $SE = 0.002$, $t(215) = 0.31$, $p = .759$. This indicated that age on its own was not driving children's exploratory behavior in this sample.

2.8 Confirmatory Analyses: Research Questions 1-3

The first three research questions relate to the main confirmatory planned analyses regarding the effect of verbal pedagogy and expected action on exploratory play. I ran a 2 x 2 factorial ANOVA for the number of unique actions outcome, and because this outcome was positively skewed (and thus violated assumptions of normality), it was log transformed prior to analysis. The proportion of time with demonstrated action outcome also didn't meet the assumptions of normality (being bimodal/U-shaped), and thus I used an arcsine square-root transformation prior to including it in a 2 x 2 factorial ANOVA (in the reporting related to research questions 1-3) in order to approximate normality and homogeneity of variance. For all additional analyses with the proportion of time with demonstrated action outcome (related to

research questions 4-7), I used a zero-one beta regression when there were mixed effects or only continuous effects in the model. In these cases, I used the zero-one beta regression approach rather than transforming the proportion of time with demonstrated action outcome since a beta distribution accommodates skewness and heteroscedasticity while the zero-one inflated beta model additionally accommodates exact 0 and 1 values which were present in these data. However, in an effort to ensure that this pre-registered analytic approach accorded with previous findings in the literature (which tend to use linear contrasts) as well as non-parametric approaches (which don't assume normal data), I undertook additional analyses related to research questions 1-3—the results of which were consistent with the findings reported below. Details of these analyses can be found in the supplementary materials: https://osf.io/h6wat/files/osfstorage. Tables 2.5, 2.6, 2.7, and 2.8 below specify the full ANOVA results and pairwise comparisons; for the main in-text reporting for research questions 1-3, I relied on reporting means, standard deviations, and the F-statistic and p-value for each finding.

2.8.1 ***Research Question 1****: To what extent will children's responses to a knowledgeable versus naïve verbal pedagogical introduction shape children's exploratory actions? To what degree will these exploratory play patterns replicate and extend the 'double edged sword of verbal pedagogy' findings?*

Against predictions, verbal pedagogy group was not a significant predictor of the proportion of time participants spent with the demonstrated action, $F(1, 213) = 0.80$, $p = .371$ (Table 2.5). However, aligned with prior findings (e.g., Bonawitz, et al., 2011), verbal pedagogy did impact the performance of unique actions (beyond presses and waves); as expected, children in the *naïve verbal pedagogy* group performed a significantly greater number of unique actions

(*M* = 0.67, *SD* = 0.09) than those in the *knowledgeable verbal pedagogy* group (*M* = 0.38, *SD* = 0.06), $F(1, 213) = 8.26$, $p < .01$ (Table 2.6).

Table 2.5. *Full ANOVA results for the* ***proportion of time with demonstrated action outcome****.*

	df	*Sum Sq*	*Mean Sq*	*F value*	*partial* η^2	η^2 *90% CI*	*p*
Verbal Pedagogy	1	0.13	0.13	0.80	0.00	[0.00, 0.03]	.371
Expected Action	1	36.12	36.12	226.45	0.52	[0.44, 0.58]	< .001
Verbal Pedagogy × Expected Action	1	0.00	0.00	0.02	0.00	[0.00, 0.01]	.877
Residuals	213	33.97	0.16				

Table 2.6. *Full ANOVA results for the number of* ***unique actions outcome****.*

	df	*Sum Sq*	*Mean Sq*	*F value*	*partial* η^2	η^2 *90% CI*	*p*
Verbal Pedagogy	1	1.52	1.52	8.26	0.04	[0.01, 0.09]	.004
Expected Action	1	1.83	1.83	9.95	0.04	[0.01, 0.10]	.002
Verbal Pedagogy × Expected Action	1	0.74	0.74	4.04	0.02	[0.00, 0.06]	.046
Residuals	213	39.23	0.18				

2.8.2 Research Question 1 Summary:

These findings indicate that *naïve verbal pedagogy* was associated with enhanced exploration, but verbal pedagogy group did not affect proportional play. The latter finding failed to replicate the anticipated main effect of verbal pedagogy on proportional play behavior (i.e., contrary to previous findings, children who heard knowledgeable statements did not spend a greater proportion of their time with the introduced action). However, current findings did replicate the anticipated effect of verbal pedagogy on unique actions (i.e., children who heard naïve statements performed a greater number of unique actions).

2.8.3 *Research Question 2:* *To what extent will action expectations shape the exploratory actions children display?*

The proportion of time with demonstrated action outcome indicates the proportion of time participants spent pressing (in the *consistent expected action* condition) or waving (in the

violating expected action condition). I did not pre-register a prediction related to the main effect of expected action group on the proportion of time spent with the demonstrated action beyond the prediction related to preliminary analyses: that those in the *consistent expected action group* would perform fewer wave actions relative to those in the *violating expected action group* (which accorded with the preliminary findings). However, I pre-registered a prediction that those who saw the *consistent expected action* would perform fewer unique actions than those who saw the *violating expected action*. As it turned out, expected action condition had an impact on both the proportion of time children spent performing the demonstrated action ($F(1, 213) = 226.45$, $p < .001$) (Table 2.5 above), and the number of unique actions they performed ($F(1, 213) = 9.95$, $p < .01$) (Table 2.6 above). Regarding the proportion of demonstrated action measure, those in the *consistent expected action* group performed a significantly greater proportion of presses (*Mprop* = 0.83, *SD* = 0.19) than the proportion of waves performed by those in the *violating expected action* group (*Mprop* = 0.24, *SD* = 0.35). Additionally, those who saw a violating action (wave) performed a significantly greater number of unique actions (*M* = 0.68, *SD* = 0.09) than those who saw a consistent (press) expected action (*M* = 0.38, *SD* = 0.07).

2.8.4 Research Question 2 Summary:

I did not have a pre-registered prediction regarding the impact of expected action group on performing unique actions; however, those who saw the expectation-consistent action spent a greater proportion of their time pressing than those who saw the expectation-violating action spent waving. As predicted in the pre-registration, viewing the expectation-violating action led children to perform a greater number of unique actions (and thus, to be more exploratory in their play behavior).

2.8.5 *Research Question 3:* *To what extent will action expectations and verbal pedagogy interact in their impact on children's exploratory play?*

For the interaction between action expectations and verbal pedagogy, I pre-registered a prediction that those in the consistent expected action group would spend a greater proportion of time pressing—especially after hearing knowledgeable verbal pedagogy—than those in the violating expected action group would spend waving. I did not pre-register a prediction regarding how the interaction between verbal pedagogy and expected action group would affect the number of unique actions performed. For the proportion of time with demonstrated action measure, no interaction between verbal pedagogy and expected action emerged ($p = .877$) (Table 2.5 above and Table 2.7 below). For the unique actions measure, however, a factorial ANOVA including both verbal pedagogy and expected action revealed a significant interaction between them ($F(1, 213) = 4.04$, $p < .05$) (Table 2.6 above). Specifically, those in the *naïve verbal pedagogy* x *violating expected action* group performed a significantly greater number of unique actions than any other group ($ps < .01$) (the rest of which did not differ significantly from one another in their number of unique actions, $ps > .852$) (Table 2.8 and Figure 2.6).

Table 2.7. *Tukey HSD* ***pairwise comparisons for the proportion of time with demonstrated action outcome*** *2x2 factorial ANOVA.*

Comparison	**Difference**	*Lower CL*	*Upper CL*	*SE*	*p (adj)*
Consistent: Knowledgeable vs. Consistent: Naïve	-0.04	-0.23	0.16	0.08	.958
Consistent: Knowledgeable vs. Violating: Knowledgeable	0.81	0.61	1.01	0.08	< .001
Consistent: Knowledgeable vs. Violating: Naïve	0.79	0.59	0.99	0.08	< .001
Consistent: Naïve vs. Violating: Knowledgeable	0.85	0.65	1.04	0.08	< .001
Consistent: Naïve vs. Violating: Naïve	0.82	0.63	1.02	0.08	< .001
Violating: Knowledgeable vs. Violating: Naïve	-0.02	-0.22	0.18	0.08	.993

Table 2.8. *Tukey HSD* ***pairwise comparisons for the number of unique actions*** *2x2 factorial ANOVA.*

Comparison	**Difference**	*Lower CL*	*Upper CL*	*SE*	*p (adj)*
Consistent: Knowledgeable vs. Consistent: Naïve	-0.06	-0.27	0.15	0.08	.896
Consistent: Knowledgeable vs. Violating: Knowledgeable	-0.07	-0.28	0.15	0.08	.853
Consistent: Knowledgeable vs. Violating: Naïve	-0.36	-0.57	-0.14	0.08	< .001
Consistent: Naïve vs. Violating: Knowledgeable	-0.01	-0.22	0.20	0.08	1.000
Consistent: Naïve vs. Violating: Naïve	-0.30	-0.51	-0.09	0.08	.002
Violating: Knowledgeable vs. Violating: Naïve	-0.29	-0.51	-0.08	0.08	.003

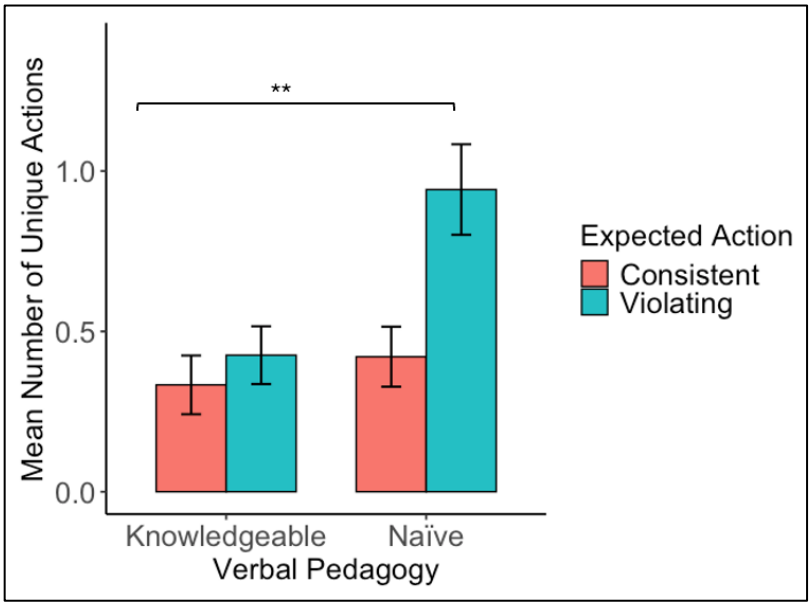


*Figure 2.6. The mean number of unique actions performed by expected action and verbal pedagogy group, with error bars. '**' indicates p < . 01 for all pairwise comparisons.*

2.8.6 Research Question 3 Summary

Taken together, the findings related to research questions 3 suggest that action expectation group and verbal pedagogy group did not interact to influence the proportion of time children spent with the demonstrated action, counter to my pre-registered prediction. However, observing an action that violated expectations following a pedagogically naïve introduction was especially powerful in enhancing children's exploration via discovering unique actions relative to any other combination of verbal pedagogy and action demonstration.

2.9 Exploratory Analyses: Research Questions 4-8

I pre-registered an exploratory analytic approach related to the variables in research questions 4-7 (EF, Effortful Control, Extraversion/Surgency), and the approach in research question 8 was also exploratory (though not pre-registered).

2.9.1 ***Research Question 4:*** *To what degree will children's executive function (EF) skills display significant associations with their patterns of exploratory play when controlling for age?*

EF and proportion of time with demonstrated action

For the proportion of time variable, a zero-one beta regression relating EF to proportion of time with the demonstrated action while controlling for age revealed that EF on its own was not associated with children's proportional play behavior, β = -0.01, *SE* = 0.01, z = -1.38, p = .166.

EF and number of unique actions. On its own, EF was not related to the number of unique actions children performed. The fixed effect of EF from a linear regression predicting the number of unique actions performed by EF score while controlling for age also failed to reach significance, β = -0.002, *SE* = 0.003, z = -0.74, p = .460.

EF and any wave attempt (violating expected action group only). I performed a further exploratory analysis looking only at the *violating expected-action* group, as it was of interest to learn the extent to which EF and/or age might have influenced whether children even considered executing the expectation-violating wave action at all in their exploratory play with the light. A logistic regression assessing whether children performed any waves with EF and age as fixed effects plus their interaction showed that while there was no fixed effect of either age (p = .488) or EF (p = .631), the interaction between age and EF was significant, β = 0.07, *SE* = 0.02, z = 3.25, p < .05. Specifically, for each month increase in age, the odds of attempting any waves in

the *violating expected action* group increased by roughly 8% for each point increase in EF (HTKS score) (Odds Ratio = 1.08, 95% CI [1.03, 1.14]).

EF and first action (violating expected action group only). I undertook another exploratory analysis to discover the extent to which EF and/or age might have been related to whether children who viewed the violating wave action went on to enact that violating wave action first, before undertaking any other action on the light. Looking only at the *violating expected action* group and focusing on whether children performed a wave as their first action, I ran a logistic regression with both EF and age as fixed effects plus their interaction. Results indicated that while neither age (p = .810) nor EF (p = .447) generated a fixed effect, there was a significant interaction between age and EF in terms of the likelihood that children performed the violating wave action first, $\beta = 0.04$, $SE = 0.02$, $z = 2.18$, $p < .05$. This interaction (see Figure 2.7) indicated that younger children were less likely to attempt a wave than older children regardless of EF score, while for older children, having higher EF scores made a difference in whether they waved as their first action during the play session (with higher EF individuals showing a higher likelihood of waving first). Specifically, for each month increase in age, the odds of attempting a wave as the first action while in the *violating expected action* group increased by about 5% for each point increase in EF (HTKS score) (Odds Ratio = 1.05, 95% CI [1.01, 1.09]).

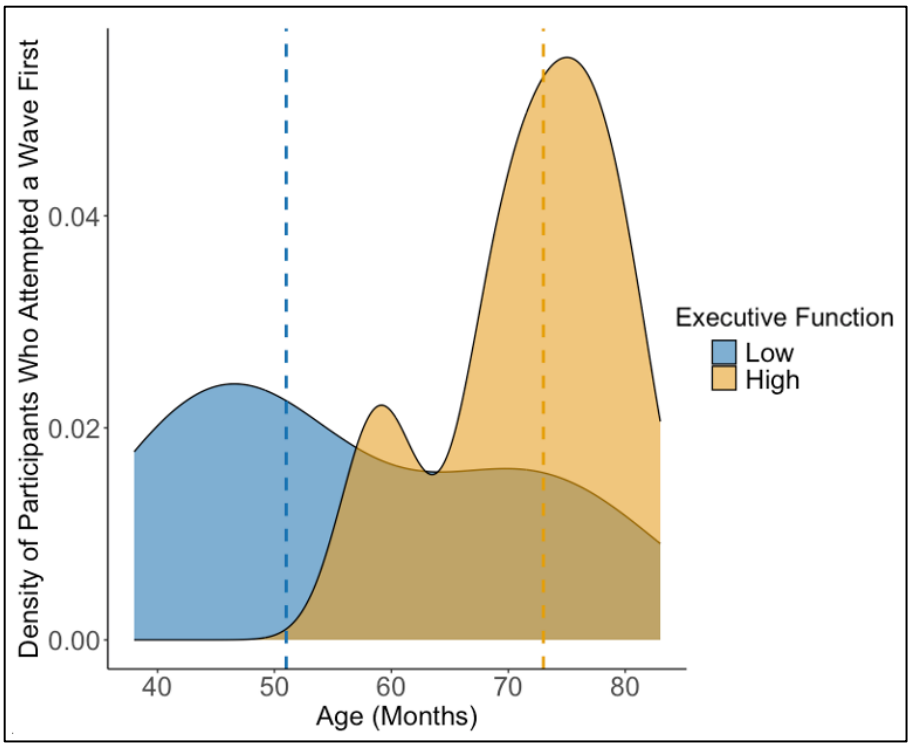


Figure 2.7. Density plot showing participants who attempted a wave first by age and EF (only data from the violated expected action condition is included in the figure. While EF was treated as continuous in the reported findings, it was grouped into high and low groups for the figure. The HTKS scores ranged from 0-52, and the cutoff of 26 distinguished participants who scored well on training and Part 1 of the task and moved on to Part 2 (designated High EF) from those who scored poorly on Part 1 or failed to complete it (designated Low EF). Vertical dotted lines represent the median age value for each EF group.

2.9.2 Research Question 4 Summary

In reference to research question 4, executive function did not show a significant effect on children's exploratory play behaviors on its own when controlling for age. However, when looking only at the violating expected action group, higher EF alongside increases in age were associated with greater odds of a) waving at all, and b) attempting a wave as the first action.

2.9.3 *Research Question 5:* *To what degree will children's EF skills interact with verbal pedagogy and action expectations in their association with exploratory play patterns when controlling for age?*

To assess the interaction between EF and the other independent variables in affecting the proportion of time with demonstrated action, I fit a zero-one beta regression to accommodate a

mixed effects approach with the proportion of time with demonstrated action outcome: I included verbal pedagogy, expected action, and EF as fixed effects (with age as a covariate) plus all of their interactions. As anticipated, there was no significant fixed effect of verbal pedagogy, but there was a significant fixed effect of expected action group ($p < .001$, Table 2.9). EF did not present a significant fixed effect on its own ($p = .268$), but an interaction between EF and expected action group emerged, $\beta = 0.07$, $SE = 0.01$, $z = -4.31$, $p < .001$. A series of planned comparisons clarified that the *violating expected action* condition was the primary locus of this interaction (see Table 2.10): In that condition, children with higher EF scores were more likely to spend a greater proportion of their time waving ($M = 0.39$, $SD = 0.39$) than those with lower EF scores ($M = 0.09$, $SD = 0.21$) (Figure 2.8).

Table 2.9. *Zero-one beta regression output with expected action, verbal pedagogy, EF (HTKS score) and age as a covariate predicting the* ***proportion of time with demonstrated action outcome.***

	Estimate	*SE*	*z*	*95%CI Lower*	*95%CI Lower*	*p*
Intercept	1.50	0.37	4.08	0.78	2.21	< .001
Expected Action	-3.44	0.51	-6.71	-4.45	-2.44	< .001
Verbal Pedagogy	-0.18	0.41	-0.45	-0.98	0.61	.655
EF (HTKS score)	-0.01	0.01	-1.11	-0.04	0.01	.268
Age	0.18	0.13	1.42	-0.07	0.43	.157
Expected Action x Verbal Pedagogy	0.86	0.66	1.29	-0.45	2.16	.197
Expected Action x EF (HTKS score)	0.06	0.01	4.31	0.03	0.09	< .001
Verbal Pedagogy x EF (HTKS score)	0.01	0.01	0.90	-0.01	0.04	.369
Verbal Pedagogy x Expected Action x EF (HTKS score)	-0.04	0.02	-2.05	-0.08	0.00	.040

Table 2.10. *Bonferroni-corrected contrasts showing the difference in EF (HTKS) slopes across expected action groups, highlighting their* ***two-way interaction*** *related to the* ***proportion of time with demonstrated action outcome.***

	Estimate	*SE*	*z*	*p (Bonf.)*
EF (HTKS) slope: Consistent vs. Violating	-0.04	0.01	-4.38	< .001

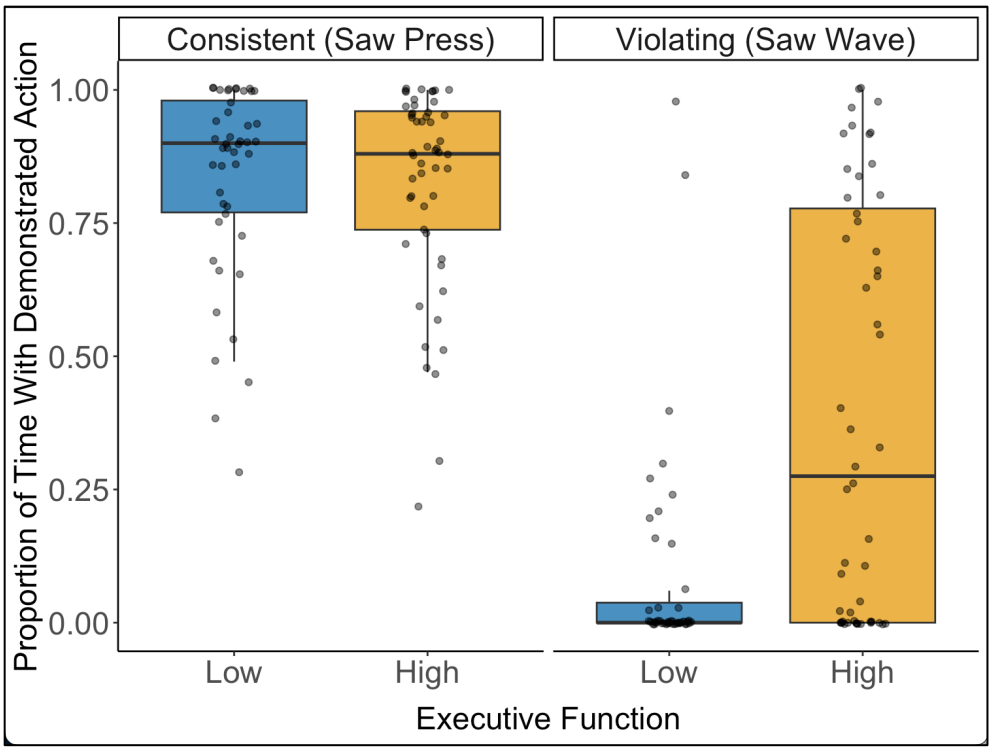


Figure 2.8. A boxplot showing the proportion of time with demonstrated action by expected action group and EF. HTKS scores range from 0-52, with the cutoff of 26 to distinguish participants who scored well on training and Part 1 of the task and moved on to to Part 2 (designated High EF) from those who scored poorly on Part 1 or failed to complete it (designated Low EF).

An additional three-way interaction emerged between expected action, EF, and verbal pedagogy, β = -0.04, *SE* = 0.02, z = -2.05, $p < .05$ (Table 2.9 above). This interaction was driven by the fact that those in the *violating expected action* group who heard *knowledgeable verbal pedagogy* showed a greater difference in their proportion of time with the demonstrated action by EF score than those in the *violating expected action* group who heard *naïve verbal pedagogy* (Tables 2.11 and 2.12, and Figure 2.9), although the Bonferroni-corrected contrast related to this interaction no longer indicated statistical significance, p = .057 (Table 2.11).

Table 2.11. *Bonferroni-corrected contrasts showing the difference in EF (HTKS) slopes across verbal pedagogy group for the violating expected action group only, highlighting their* ***three-way interaction*** *related to the* ***proportion of time with demonstrated action outcome for one level (violating).***

	Estimate	*SE*	*z*	*p (Bonf.)*
Knowledgeable vs. Naive	0.03	0.01	1.90	.057

Table 2.12. *Bonferroni-corrected pairwise comparison showing the difference in EF (HTKS) slopes across verbal pedagogy groups for the consistent expected action group only, highlighting their* ***three-way interaction*** *related to the* ***proportion of time with demonstrated action outcome for one level (consistent).***

	Estimate	*SE*	*z*	*p (Bonf.)*
Knowledgeable vs. Naive	-0.01	0.01	-0.90	.369

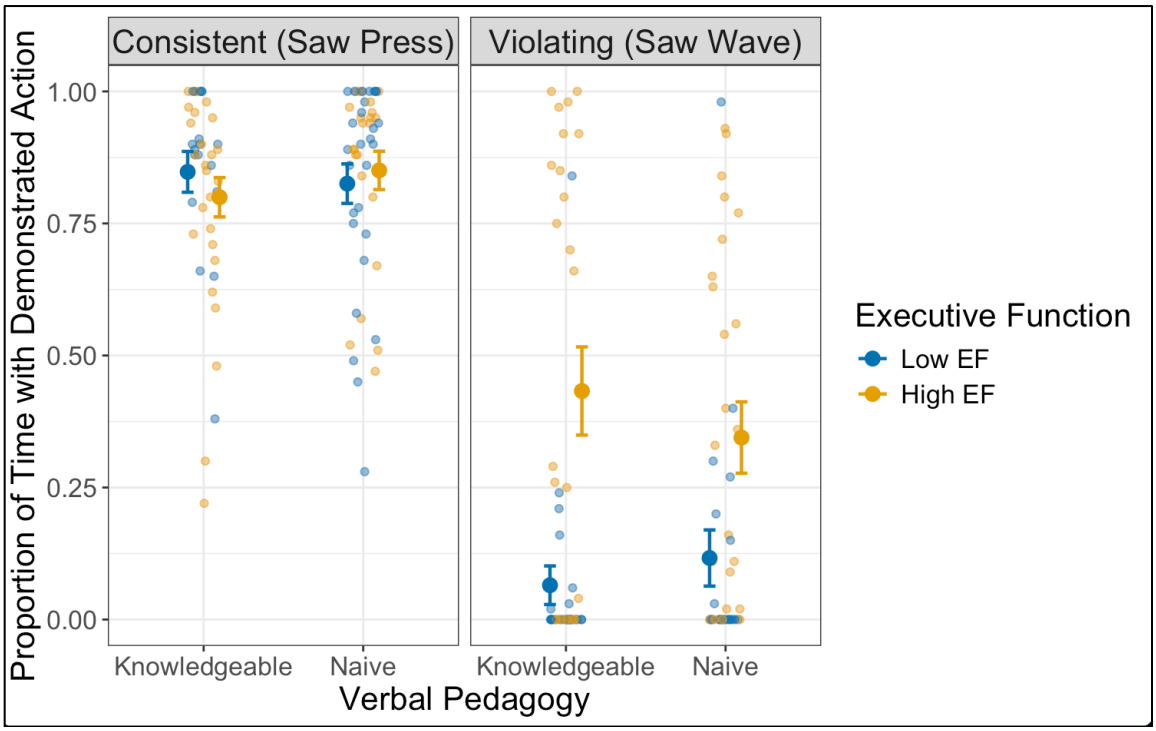


Figure 2.9. A comparison of proportion of time with demonstrated action by verbal pedagogy, visual affordance, and EF group, with error bars. The HTKS scores range from 0-52, with the cutoff of 26 to distinguish participants who scored well on training and Part 1 of the task and moved on to Part 2 (designated High EF) from those who scored poorly on Part 1 or failed to complete it (designated Low EF).

I was additionally interested in the extent to which EF impacted children's pressing behavior in the *violating expected action* group. The focus of this analysis concerned the proportion of time spent with the non-demonstrated action (rather than the demonstrated action, as in prior analyses). I fit another zero-one beta regression with pedagogy, expected action and EF as fixed effects (with age as a covariate) plus all their interactions in relation to the proportion of time spent with the non-demonstrated action. All fixed and interaction effects were significant (Table 2.13). Focusing on relations to EF, Table 2.14 shows that those in the *violating expected*

action group did not vary in their pressing behavior by EF. Table 2.15 shows that those in the *consistent expected action* group did vary in their waving behavior by EF. This indicates that EF apparently did not show a relationship with children's ability to suppress a prepotent action (pressing) in the *violating expected a*ction group. However, EF did show an association with children's likelihood of discovering the violating action (waving) in the *consistent expected action* group.

Table 2.13. *Zero-one beta regression output with expected action, verbal pedagogy, EF (HTKS score) and age as a covariate predicting the* ***proportion of time with non-demonstrated action outcome.***

	Estimate	*SE*	*z*	95%*CI Lower*	95%*CI Lower*	*p*
Intercept	-15.26	5.32	-2.87	-25.69	-4.83	.004
Expected Action	16.06	5.31	3.02	5.65	26.46	.002
Verbal Pedagogy	10.64	5.38	1.98	0.11	21.18	.048
EF (HTKS score)	0.32	0.12	2.70	0.09	0.54	.007
Age	-0.47	0.16	-3.03	-0.78	-0.17	.002
Expected Action x Verbal Pedagogy	-10.71	5.38	-1.99	-21.26	-0.16	.047
Expected Action x EF (HTKS score)	-0.32	0.12	-2.79	-0.55	-0.10	.005
Verbal Pedagogy x EF (HTKS score)	-0.24	0.12	-2.05	-0.47	-0.01	.041
Expected Action x Verbal Pedagogy x EF (HTKS score)	0.24	0.12	2.03	0.01	0.47	.043

Table 2.14. *Bonferroni-corrected contrasts showing the difference in verbal pedagogy estimates by EF (HTKS) group (-1 SD, mean, 1 SD) for the* ***violating expected action*** *group only, highlighting their* ***three-way interaction*** *related to the* ***proportion of time with non-demonstrated action outcome.***

EF Level	**Contrast**	**Estimate**	*SE*	*z*	*p (Bonf.)*
-1 SD	Knowledgeable vs. Naive	0.06	0.43	0.15	.883
Mean	Knowledgeable vs. Naive	0.06	0.42	0.15	.877
+1 SD	Knowledgeable vs. Naive	0.07	0.41	0.16	.870

Table 2.15. *Bonferroni-corrected contrasts showing the difference in verbal pedagogy estimates by EF (HTKS) group (-1 SD, mean, 1 SD) for the* ***consistent expected action*** *group only, highlighting their* ***three-way interaction*** *related to the* ***proportion of time with non-demonstrated action outcome.***

EF Level	Contrast	Estimate	SE	z	$p\,(Bonf.)$
-1 SD	Knowledgeable vs. Naive	-10.89	5.49	-1.98	.048
Mean	Knowledgeable vs. Naive	-10.64	5.38	-1.98	.048
+1 SD	Knowledgeable vs. Naive	-10.40	5.26	-1.98	.048

Turning to the unique actions measure, a linear regression was fit containing verbal pedagogy, expected action group, and EF as fixed effects plus their interactions (with age as a covariate). As in the ANOVA findings, there were main effects of expected action and verbal pedagogy. There was also a significant three-way interaction between EF, verbal pedagogy, and expected action group in association with the (log-transformed) number of unique actions participants performed, $\beta = 0.001$, $SE = 0.004$, $t(184) = 1.97$, $p < .05$ (see Figure 2.10 and Table 2.16, where the *p*-value rounded to .05 for this interaction, despite being slightly less than .05). To examine this interaction, I first looked only at the *violating expected action* group. Children high in EF who heard *naïve verbal pedagogy* performed a greater number of unique actions ($M = 1.19$, $SD = 1.10$) than those high in EF who heard *knowledgeable verbal pedagogy* ($M = 0.35$, $SD = 0.56$) (though the Bonferroni-corrected comparison examining this three-way interaction in Table 2.18 was no longer significant ($p = .07$)). The *consistent expected action* groups showed no significant differences in their unique actions by EF score or *verbal pedagogy* ($p = .323$) (Table 2.17).

Table 2.16. *Linear regression output with expected action, verbal pedagogy, EF (HTKS score) and age as a covariate predicting the* ***number of unique actions outcome.***

	Estimate	*SE*	*t*	*95%CI Lower*	*95%CI Lower*	*p*
Intercept	0.33	0.03	10.30	0.26	0.39	< .001
Expected Action	-0.10	0.03	-3.35	-0.17	-0.04	< .001
Verbal Pedagogy	-0.10	0.03	-3.05	-0.16	-0.03	.003
EF (HTKS score)	0.00	0.00	-0.39	-0.01	0.00	.694
Age	0.00	0.00	0.39	-0.01	0.01	.697
Expected Action x Verbal Pedagogy	0.06	0.03	1.92	0.00	0.12	.057
Expected Action x EF (HTKS score)	0.00	0.00	-0.32	0.00	0.00	.753
Verbal Pedagogy x EF (HTKS score)	0.00	0.00	-0.50	0.00	0.00	.619
Expected Action x Verbal Pedagogy x EF (HTKS score)	0.00	0.00	1.97	0.00	0.01	.050

Table 2.17. *Bonferroni-corrected contrasts showing the difference in EF (HTKS) slopes across verbal pedagogy groups for the consistent expected action group only, highlighting their* ***three-way interaction*** *related to the* ***number of unique actions outcome for one level (consistent).***

Contrast	**Estimate**	*SE*	*t*	$p\,(Bonf.)$
Knowledgeable vs. Naive	0.01	0.01	0.99	.323

Table 2.18. *Bonferroni-corrected contrasts showing the difference in EF (HTKS) slopes across verbal pedagogy groups for the violating expected action group only, highlighting their* ***three-way interaction*** *related to the* ***number of unique actions outcome for one level (violating).***

Contrast	**Estimate**	*SE*	*t*	$p\,(Bonf.)$
Knowledgeable vs. Naive	-0.01	0.00	-1.82	.070

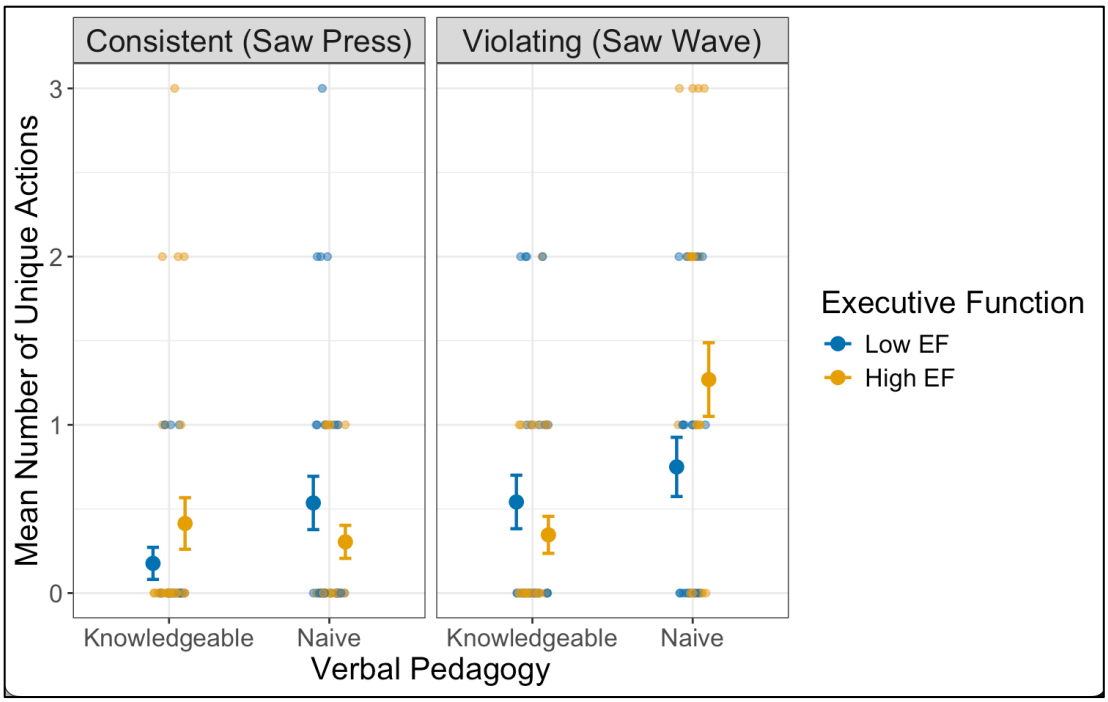


Figure 2.10. A comparison of unique actions by verbal pedagogy, visual affordance, and EF group, with error bars. The HTKS scores range from 0-52, with the cutoff of 26 to distinguish participants who scored well on training and Part 1 of the task and moved on to Part 2 (designated High EF) from those who scored poorly on Part 1 or failed to complete it (designated Low EF).

2.9.4 Research Question 5 Summary

In reference to research question 5, EF interacted with expected action to predict the proportion of time children spent with the demonstrated action, as those in the *violating expected action* group showed a greater likelihood of waving if they had high EF scores while the proportion of pressing did not differ by EF in the *consistent expected action* group. There was also a three-way interaction between EF, expected action, and verbal pedagogy, as those in the *violating expected action* group who had high EF and heard *knowledgeable verbal pedagogy* waved more than those who were low in EF and heard *knowledgeable verbal pedagogy*, a difference in proportional play behavior that was impacted by EF score more than for other groups (*naïve x violating, knowledgeable x consistent, naïve x consistent*). However, this was a marginal effect which failed to retain significance following the Bonferroni correction. EF also

influenced whether those in the *consistent expected action* group discovered waving as a secondary action, with those higher in EF more likely to make this discovery.

In relation to the number of unique actions, there was a three-way interaction between EF, expected action, and verbal pedagogy. Those in the *violating expected action* group who heard *naïve verbal pedagogy* showed a difference in the number of unique actions by EF scores (though again, this was only a marginal effect). Specifically, the higher children's HTKS scores, the greater the number of unique actions they tended to display, indicating intensified exploratory activity. This difference in unique actions by EF scores was not observed in any other group, pointing to children higher in EF showing an especially strong exploratory response to the combination of naïve pedagogy and an expectation violation.

2.9.5 *Research Question 6:* *To what extent will parental responses to the CBQ scales Effortful Control and Extraversion/Surgency show significant relationships to children's exploratory play behaviors?*

From the CBQ, two scales were created from subscale data: Extraversion/Surgency (subscales: Impulsivity, Activity Level, Approach, (loading negatively) Shyness) and Effortful Control (subscales: Attentional Focusing and Inhibitory Control). Though I initially considered controlling for age in relation to Effortful Control and Extraversion/Surgency analyses, age was not related to Extraversion/Surgency ($r(198) = -.05$, $p = .463$), and only weakly related to Effortful Control ($r(198) = .25$, $p < .001$). Therefore, I opted to leave age out as a covariate in the analyses related to research questions 6 and 7. Both unique actions and proportion of time with demonstrated function outcomes were included in separate models (linear regression in the case of the unique actions outcome and zero-one beta regression in the case of the proportion of time with demonstrated action). These models contained Effortful Control and Extraversion/Surgency

as fixed effects plus their interaction while controlling for age, and all were significant for the proportion of time outcome (but not for the unique actions outcome) (see Tables 2.19 and 2.20 for coefficients, p-values, and contrasts, and Figure 2.11 showcasing the interaction).

Table 2.19. *Zero-one beta regression output with Effortful Control and Extraversion/Surgency predicting the* ***proportion of time with demonstrated action outcome.***

	Estimate	*SE*	*z*	95%*CI Lower*	95%*CI Lower*	*p*
Intercept	11.98	3.56	3.37	5.01	18.95	< .001
Effortful Control	-2.47	0.76	-3.24	-3.96	-0.98	.001
Extraversion/Surgency	-2.40	0.69	-3.49	-3.75	-1.05	< .001
Effortful Control x Extraversion/Surgency	0.52	0.15	3.48	0.23	0.82	< .001

Table 2.20. *Bonferroni-corrected contrasts for the zero-one beta regression with Effortful Control and Extraversion/Surgency predicting the* ***proportion of time with demonstrated action outcome****. Estimates represent the difference in slopes of Effortful Control across Extraversion/Surgency levels (-1 SD (Low), mean, +1 SD (High)) relating to* ***the two-way interaction****.*

	Estimate	*SE*	*z*	*p* (*Bonf.*)
Effortful Control slope: Low vs. Mean Extraversion	-0.52	0.15	-3.48	.001
Effortful Control slope: Low vs. High Extraversion	-1.05	0.30	-3.48	.001

2.9.6 Research Question 6 Summary

Children who scored higher in Effortful Control and those who scored higher in Extraversion/Surgency showed a tendency to spend more time with the demonstrated action. Also, for children high in Extraversion/Surgency, strong Effortful Control especially mattered for how much time they spent with the demonstrated action (the higher the Effortful Control, the greater the proportion of time they spent with the demonstrated action).

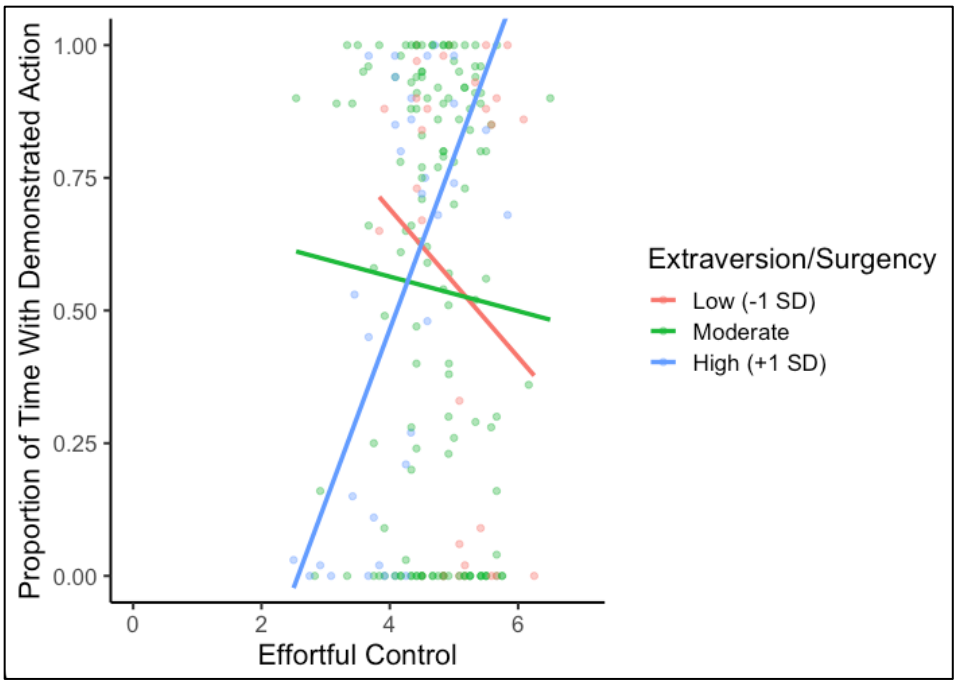


Figure 2.11. Proportion of time with the demonstrated action by Effortful Control and Extraversion-/Surgency group (low (-1 SD), moderate (mean), and high (+1 SD))—note: Extraversion/Surgency was treated as continuous in the main zero-one beta regression model. Children's scores on the parent-reported Effortful Control variable ranged from 2.5-6.5 (possible range: 0-7).

2.9.7 *Research Question 7:* *To what degree will CBQ scales Effortful Control and Extraversion/Surgency interact with verbal pedagogy and action expectations in their relationship to children's exploratory play?*

Extraversion/Surgency and Effortful Control showed no fixed or interaction effects when both verbal pedagogy and expected action were included in regressions assessing their effect on unique actions and proportion of time with demonstrated action. However, for the proportion of demonstrated action outcome (but not for the unique action outcome in a linear regression), when Effortful Control and Extraversion/Surgency were included in a zero-one beta regression with expected action group only, there was a significant three-way interaction effect, $\beta = 0.60$, $SE = 0.29$, $z = 2.08$, $p < .05$) (see Tables 2.21, 2.22, and 2.23, and Figure 2.12).

Table 2.21. *Zero-one beta regression output with Effortful Control, Extraversion/Surgency, and Expected Action predicting the* ***proportion of time with demonstrated action outcome****.*

	Estimate	*SE*	*z*	95%*CI Lower*	95%*CI Lower*	*p*
Intercept	-1.25	5.32	-0.23	-11.67	9.17	.815
Expected Action	9.70	6.74	1.44	-3.51	22.90	.150
Effortful Control	0.57	1.13	0.50	-1.65	2.79	.615
Extraversion/Surgency	0.51	1.07	0.48	-1.59	2.62	.634
Expected Action x Effortful Control	-2.70	1.44	-1.88	-5.51	0.12	.061
Expected Action x Extraversion/Surgency	-2.49	1.33	-1.87	-5.10	0.11	.061
Effortful Control x Extraversion/Surgency	-0.12	0.23	-0.51	-0.57	0.33	.608
Expected Action x Effortful Control x Extraversion/Surgency	0.60	0.29	2.08	0.04	1.17	.037

Table 2.22. *Bonferroni-corrected contrasts for the zero-one beta regression with Effortful Control and Extraversion/Surgency predicting the* ***proportion of time with demonstrated action outcome****. Estimates represent the difference in slopes of Effortful Control across Extraversion/Surgency levels (-1 SD (Low), mean, +1 SD (High)) related to the* ***three-way interaction (at the violating level for expected action group).***

	Estimate	*SE*	*z*	*p (Bonf.)*
Effortful Control slope: Low vs. Mean Extraversion	-0.48	0.17	-2.77	.017
Effortful Control slope: Low vs. High Extraversion	-0.97	0.35	-2.77	.017

Table 2.23. *Bonferroni-corrected contrasts for the zero-one beta regression with Effortful Control and Extraversion/Surgency predicting the* ***proportion of time with demonstrated action outcome****. Estimates represent the difference in slopes of Effortful Control across Extraversion/Surgency levels (-1 SD (Low), mean, +1 SD (High)) related to the* ***three-way interaction (at the consistent level for expected action group).***

	Estimate	*SE*	*z*	*p (Bonf.)*
Effortful Control slope: Low vs. Mean Extraversion	0.12	0.23	0.51	1.000
Effortful Control slope: Low vs. High Extraversion	0.24	0.46	0.51	1.000

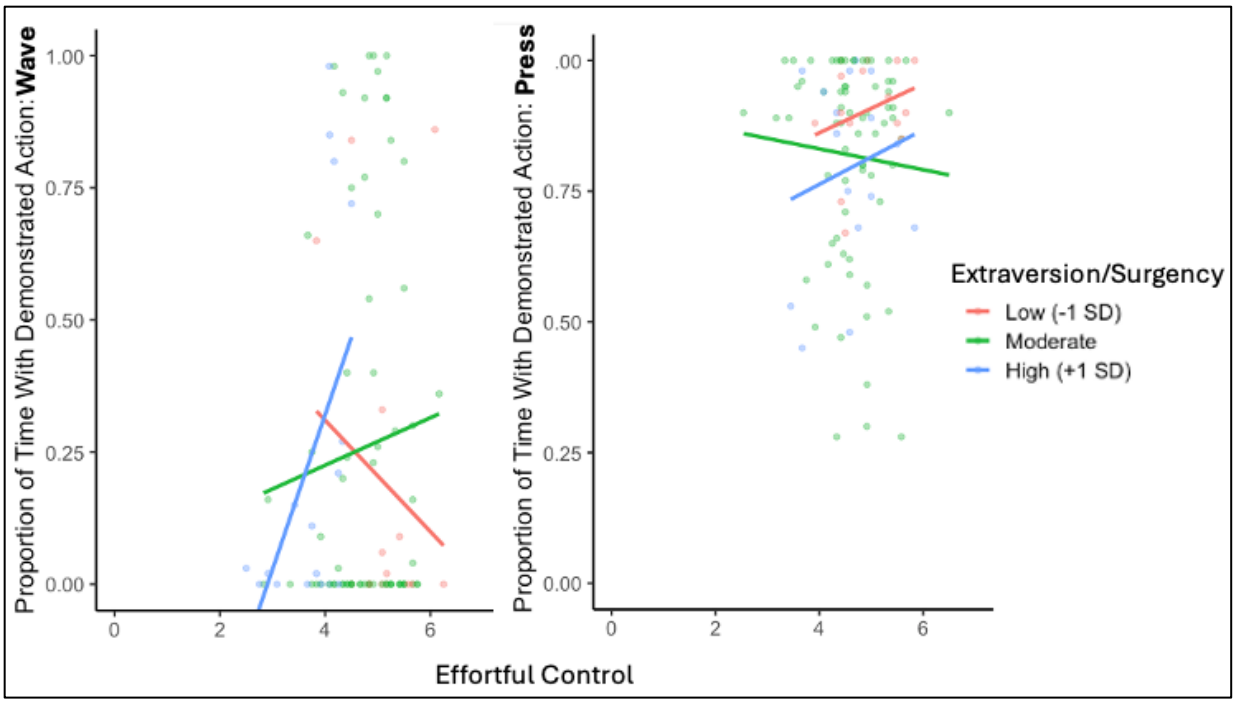


Figure 2.12. Proportion of time with the demonstrated action by expected action group, Effortful Control, and Extraversion/Surgency group (low (-1 SD), moderate (mean), and high (+1 SD)).

2.9.8 Research Question 7 Summary

Extraversion/Surgency and Effortful Control did not interact with verbal pedagogy to influence children's exploratory play, but they showed a three-way interaction with expected action group, such that those high in Extraversion/Surgency and high in Effortful Control were especially focused on waving in the *violating expected action* group. For those in the *consistent expected action* group, Effortful Control and Extraversion/Surgency were not associated with pressing activity.

2.9.9 *Further (non a priori) Exploratory Research Question 8*: *To what extent was pressing an exploratory behavior for those in the violating expected action group who later went on to wave? Is this pattern of behavior related to the other play finding having to do with enhanced exploration (number of unique actions)?*

In another exploratory analysis, I was additionally interested in the portion of participants who first pressed but later waved in the *violating expected action* group. In particular, I was

interested in whether pressing might operate as a type of exploratory behavior for this group; that is, participants might first press (rather than wave) if they are curious if the light can be turned on in the expected manner (a press) as well as in the unexpected manner (wave) that they had just observed. If participants first pressed and later waved out of exploratory curiosity, they might attempt a greater number of unique actions than those who first pressed but never waved. This turned out to be the case, as those in the *violating expected action* group who first pressed and later waved performed a greater number of unique actions (M = 1.07, SD = 1.16) than those who first pressed but never waved (M = 0.49, SD = 0.79), $F(1, 56) = 4.79$, $\eta p^2 = .08$, 90% CI [0.00 , 0.20], $p < .05$.

2.9.10 Research question 8 Summary

Children in the *violating expected action* group who first pressed and later waved performed a significantly greater number of unique actions than those who first pressed but never waved. This finding suggests that first enacting a press of the light after having witnessed a wave may have reflected exploratory behavior on children's part, rather than simply indicating a prepotent tendency to perform the action the light afforded (or a failure to have even noticed the waving action in the first place).

Discussion

In Study 1, I was interested primarily in the effect of verbal pedagogy and action expectations on children's exploratory play, and secondarily on the effect these variables on exploratory play when taking EF, age, and temperament into account. Results initially revealed that pressing was a highly salient action for this light across conditions, confirming that children indeed expected a pressing action to turn on the light. Other *a priori* predictions I pre-registered were also confirmed: whether children saw a press or a wave in the slideshows strongly

predicted the proportion of time they spent with the introduced action. Those who saw an action consistent with their prior expectations (press) were especially likely to focus on pressing, while those who saw an action violating their prior expectations (wave) were likely to perform both presses and waves. That said, even this group performed a greater proportion of presses than waves, speaking again to children having strong expectations that the light afforded pressing (which was by design).

Pedagogical context (knowledgeable vs. naïve) did not impact children's proportional play behavior overall. This result did not accord with previous findings (Bonawitz et al., 2011; Zsoldos & Kiraly, 2025; Brezack et al., 2021), yet I do not regard this as a direct replication failure per se, for at least two reasons. First, pressing was clearly highly salient for children as an action relevant to turning on the light, which exerted strong influence on their play activities; thus, the salience of pressing may have rendered children less sensitive to the manipulation of pedagogical context, or overwhelmed their ability to be guided in their play activities by that context. Second, the proportional play behavior dependent measure holds somewhat different meaning in the present study relative to the prior studies in which it was utilized. In those studies, proportion time spent with the demonstrated action was understood to be indicative of suppression of exploration – the higher the proportion, the more children were constraining themselves to enacting the demonstrated action. In the present study, this interpretation of the proportional play behavior dependent variable was not straightforwardly viable. Consider, for example, the *consistent expected action* group: they strongly expected a press would turn on the light and observed a pressing action in fact turn on the light. This presents a rather dull scenario, which could have suppressed a desire to spend much time enacting the demonstrated action (pressing), regardless of pedagogical context, thereby undercutting the previously documented

effect of pedagogical context. And consider also the *violating expected action* group. This group expected that pressing would turn on the light, but what they observed instead was that the actor turned on the light by waving over it. If they indeed noticed the violation of expectation, it would seem likely to have triggered questions, such as: Why did the actor wave instead of press? How does waving work to turn on the light? Will the light turn on if I also do a waving action over the light? Did the actor wave because pressing doesn't work to turn on the light? Do both pressing and waving work? Am I supposed to wave or is it okay if I press on the light? As such questions arose for children, they may have engaged in relevant exploratory activities and they likely would have formed concomitant inferences, which presumably would have guided their subsequent exploratory activities. Some of these questions and inferences would have promoted waving, others pressing, and others both waving and pressing. And children may have entertained more than one of these questions and/or inferences over the play period. The upshot is that the *violating expected action* condition in the present study would be expected to provoke both waving and pressing, to varying degrees depending on which questions and inferences children were entertaining at any given phase of the play period. It is not clear that a knowledgeable pedagogical context would operate to constrain children's activities to the demonstrated action in the face of such inferential complexity. All in all, the lack of an overall effect of pedagogical context on children's proportional play behavior in the present study does not represent a failure to replicate prior findings in which action expectations were not under manipulation.

The findings also confirmed that those who heard naïve pedagogy and saw a violating event were especially likely to show enhanced exploration relative to other groups, as measured by their performance of unique actions. Exploratory analyses revealed that this pattern was

additionally associated with children's executive skill, as those in the *naïve verbal pedagogy*-plus-*violating expected action* group who had higher EF scores performed more unique actions relative to those who had lower EF scores.

The first action children performed was indicative of their exploration as well, especially for those who saw the violating event. In particular, those who saw the violating event and pressed as their first action, and who also went on to wave might be using pressing to explore additional causal features of the toy that they hadn't yet seen relative to those who first pressed but never waved. While this tendency was related to the other main exploratory play outcome (unique actions), future work is needed to understand how performing a salient action can relate to exploratory behavior more broadly. A secondary explanation is that children were using their first action to generate data to disambiguate the full causal structure of an object (similar to Sobel et a., 2022). However, in order to understand if this was driving children's behavior, we would need to understand more about whether all children in this group registered the violating event at all (see results in Chapter 3), and this remains an open area for future work.

To sum up Study 1, I found that 3-to-7-year-old children's exploratory activities arose as a joint product of the pedagogical features of the learning situation, their expectations for causal action, and their direct observation of another's causal action. As well, skills children brought to the learning situation mattered in relation to their exploratory activities. In particular, children with stronger executive skills were especially exploratory following a violating event, suggesting that their executive skill promoted exploratory activities that were more systematically guided by questions and inferences, or more flexible knowledge-updating processes, or both.

In the following chapter, I focus on the data I collected regarding children's attentional patterns as they experienced the unfolding causal-action scenarios already described; that is, the

consistent and violating expected action events. I will discuss evidence related to children's implicit cognitive engagement while viewing action demonstrations in order to understand more about the mechanisms supporting children's inference-updating as events unfold.

CHAPTER III:

STUDY 2: DWELL TIME AS A WINDOW INTO HOW VERBAL PEDAGOGY AND ACTION EXPECTATIONS JOINTLY SHAPE CHILDREN'S EVENT PROCESSING

Introduction

In Study 1, children showed differences in their exploratory play behavior based upon both pedagogical and visual cues during an introduction to a new object. Specifically, children explored more when they both heard naïve pedagogical statements and saw events that violated their prior expectations relative to those who heard knowledgeable pedagogy and saw events consistent with their prior expectations. Children were likely to try the action that was visually afforded by the object (an action consistent with prior expectations), even when they saw a violating event, though the likelihood of performing the violating action grew with executive function abilities. Taking findings across the two primary dependent measures together, higher executive skills turned out to be associated with greater sensitivity to pedagogical context (i.e., displaying the double-edged sword trade-off more strongly) and higher-fidelity action (e.g., waving when waving was demonstrated). As well, children with stronger executive skills were more likely to discover, through their own exploratory activities, the potential for waving to illuminate the light, suggesting that executive skills also offer an advantage in observing the effects of one's own causal actions. These findings partially replicated and extended the Bonawitz and colleagues' 'double-edged sword of pedagogy' outcomes.

Turning to Study 2, I examined the nature of children's attention as they viewed the event prior to playing with the object—information that provides a window into children's processing of the action demonstration prior to their direct exploratory engagement with the novel object. To examine children's event processing, I used a technique called the dwell-time paradigm, which offered the opportunity to observe children's attentional patterns as they watched the adult

act on the novel object after hearing either a *knowledgeable* or *naïve verbal pedagogy* statement. The dwell-time paradigm was first introduced by Hard et al. (2011) as a tool to index cognitive engagement during event processing. Dwell-time procedures typically operate by capturing still frames of a pre-recorded video at a set increment (e.g., every 2 frames or 5 frames) and allowing participants to advance through these still frames at their own pace.

Several noteworthy findings arose from Hard and colleagues' seminal research using this method, which were confirmed in research of others who subsequently adopted it (e.g., Baldwin & Kosie, 2021). First, when participants dwell through slideshows depicting familiar, well-understood events, their attention tends to surge at slides depicting boundary regions within the stream of motion, a so-called "boundary advantage." Boundary regions occur when one action transitions to the next, as when, for example, a reaching motion toward a cup culminates in a grasp and then transitions into a lifting motion. Boundary regions are thought to garner intensified attention because they offer rich causal and predictability information: as one event ends, there is uncertainty about what will happen next, making the transitional boundary region valuable for gleaning what happens next and thereby predicting how things will continue to unfold.

Dwell-time patterns documented in previous research also display evidence of the hierarchical structure viewers impose on the stream of motion as they view it. Specifically, participants tend to display a "hierarchical advantage" pattern in their dwell times: they dwell longer on slides depicting high-level "coarse-grained" boundaries – such as transitions from one major task to another (e.g., an actor moves from the kitchen to the living room on having completed kitchen clean-up) – than slides depicting more low-level, "fine-grained" boundaries – such as transitions from one small-scale motion element to another (e.g., an actor engaged in

kitchen clean-up transitions from moving a brush to make contact with a plate, to scrubbing the plate with the brush). Slides depicting content midstream between event boundaries – "non-boundary" content – tends to receive the least attention from viewers. Coarse-grained level boundaries appear to receive enhanced attention because they offer powerful causal information that propagates through a subsequent sequence of smaller, fine-grained level actions. In other words, coarse-grained boundaries support a gist-level analysis that informs the way attention is deployed for whole swaths of subsequent motion analysis.

Like adults, children also display boundary and hierarchical advantage patterns in their dwell times (Kosie & Baldwin, 2021; Meyer, et al., 2011). One might wonder if these patterns simply reflect attentional orienting to especially salient junctures within the flow of motion, such as points where physical motion change is greatest. That is not the case: boundary and hierarchical advantages remain even when controlling for slide-to-slide pixel differences, an index of physical motion change (Hard, et al., 2011; Kosie & Baldwin, 2021).

Another key finding is that dwell-time patterns such as the boundary and hierarchical advantage reflect learning. For example, viewers observing a novel sequence of small motion elements initially display a boundary advantage only to the fine-grained transitions between these elements. As viewing proceeds, however, their dwell times increasingly reflect sensitivity to coarse-grained boundaries reflecting predictable transitions between combinations of these small motion elements (Hard, et al., 2019). Achieving a causal analysis of novel event streams seems to be a key driver of such attentional reorganization (Buchsbaum, et al., 2009; Kosie & Baldwin, 2019). In particular, dwelling tends to surge when novel or expectation-violating content is first observed (Sage & Baldwin, 2014; Kosie & Baldwin, 2019); on subsequent

viewing, dwell-time patterns reveal emerging sensitivity to causally-relevant boundaries within that novel content (Kosie & Baldwin, 2019).

Given such prior findings, the dwell-time paradigm seemed well-suited to the present investigation. It offered the opportunity to address a suite of questions pertinent to the overall aim of discovering how verbal pedagogy, action expectations, and individual differences such as executive skill might jointly shape children's knowledge updating. Further, while other methods that record children's looking behaviors have been used in related literature (e.g., pupillometry, looking time) (Colantonio et al., 2023; Sim & Xu, 2018), recording children's dwell times provided a unique opportunity to capture children's attentional patterns in a museum setting where the other methods would be difficult to employ (while at the same time according with a great deal of literature in this area that utilizes samples recruited in museum settings (Corriveau et al., 2016)).

As described, introducing the dwell time paradigm made it possible to measure differences in attentional patterns that related to both the event itself (*consistent* vs. *violating*) and to the pedagogical statement that introduced the event (*knowledgeable* vs. *naïve*). Specifically, I expected that children would show increased dwelling to boundary over non-boundary regions within the consistent and violating expected-action slideshows. Given the novelty of using this procedure as a metric for children's attention within causal learning scenarios, additional hypotheses—including the impacts of verbal pedagogy and action expectations—fall within an exploratory vein.

For Study 2, I was interested in answering the following research questions related to both confirmatory and exploratory pre-registered analyses:

Confirmatory

1. To what degree will children's dwell time patterns replicate prior dwell time findings? That is, will analyses reveal a significant boundary advantage, as predicted (i.e., significantly higher dwell times during boundary regions relative to non-boundary regions) when controlling for pixel change?

Exploratory

2. To what degree will children's *a priori* expectations regarding object-related action influence their dwell-time patterns, both directly and in interaction with the pedagogical quality in which the object is introduced (i.e., knowledgeable vs. naïve conditions)?
3. To what extent will children's dwell-time patterns change from first to second viewing, and to what degree will such changes across viewings interact with pedagogical context (i.e., knowledgeable vs. naïve conditions) and action expectations (i.e., consistent versus violated expectations)?

Method

3.2 Participants

The sample was the same as in Study 1, with a final N including 220 participants. As mentioned in Chapter 2, I performed an *a priori* power analysis based upon similar research in this area which involved recording preschoolers' dwell time (Kosie & Baldwin, 2021), and our sample attained an acceptable degree of power (0.88) despite some minor constraints related to a June 2025 data collection cut-off. Exclusion criteria for the dwell time data were set by an outlier determination process that is described in the 'Data Preparation' section below.

3.3 Materials

3.3.1 Training Stimuli

Nine still images of a cartoon train were utilized for the training portion of the task. Children began by clicking the mouse, causing the images to change and giving the appearance that the train was moving across the screen from right to left. The purpose of the training was two-fold: First, it allowed participants to understand the mechanics of the task in terms of the affiliation between their clicking and seeing changes in the event. Second, given that the dwell-time method records attentional patterns related to the stimuli, it allowed researchers an opportunity to clarify to children that they needed to be looking at the screen when clicking the mouse to advance slides. In other words, it was important that participants understood that they needed to attend fully to the events as they were clicking, and the training task helped to set this standard for the remainder of the task. That said, our research team nevertheless undertook post hoc coding for look-aways during children's slide-show viewing, as recorded in the coding section below. The training images are available for viewing in the supplementary materials: https://osf.io/h6wat/files/osfstorage

3.3.2 Replication Stimuli

To assess if our sample showed expected overall dwell-time trends for this age range (including amplified attention to boundary regions), I utilized a set of images created for a similar study with a preschool-aged sample (Kosie & Baldwin, 2021). The previous study took place in a lab setting, so this replication effort also allowed an assessment of whether the museum setting introduced variance in dwell-time patterns not seen in previous lab findings for the identical set of stimuli. The images depicted a side-view of an adult sitting at a table with a barrier in front of them and a ball on the other side of the barrier. The adult reaches with their

right hand over the barrier and picks up the ball, bringing it back over the barrier and laying it down on the table in front of them. The full set of images and further descriptions can be found in the original paper and on the related Open Science Framework page for the study for which the stimuli were created: https://osf.io/m6q7r/overview

3.3.3 Main Dwell Time Stimuli

The main dwell time stimuli were introduced in the materials for Study 1 as the 'Expected Action Slideshows' (Figure 2.1, Study 1). Participants either saw an event where a red light turned on via a button-press action, or they saw an event where a red light turned on via a wave-over action.

3.3.4 Pedagogy Videos and Novel Object

Participants viewed the same pedagogy videos and played with the same novel object that was described in the materials for Study 1. Participants either heard a *knowledgeable verbal pedagogy* statement "This is how my toy works," or they heard a *naïve* statement: "I wonder how this toy works."

3.4 Design

The design of Study 2 was identical to Study 1, with the main difference being that the results for Study 2 reflect the dwell-time data that were collected during slideshow viewing. Participants were sorted into one of four equally sized, counterbalanced conditions, and saw one instance of pedagogy (*knowledgeable* or *naïve*) and one version of the expected action slideshow (*consistent*: press or *violating:* wave).

3.5 Procedure

The general procedure was identical to Study 1. Prior to starting the slideshow viewing, participants were told that they could stop at any time. When they expressed an interest in being

done with the task or wanting to leave the area early, researchers led them away from the table and communicated with their parents. I retained a participant's data on the dwell-time portion of the task if they made it through the slideshow training, replication slideshow, and the first viewing of the slideshow that followed the pedagogy video. In such cases, these participants were not included in dwell-time analyses assessing differences in first vs. second viewing.

While the consent process was taking place, a research assistant launched the PsychoPy® project from a laptop on the table so that it was ready once the participant was seated. When participants were first seated at the table, the researcher had them put on noise-cancelling headphones (the researcher sat next to the participant, who could still hear them through the headphones). The screen showed a sky with clouds and the researcher read from it: "First, we're going to play a game on the computer. Click on the mouse to change the pictures." The child was encouraged to click, and the same photo with clouds appeared again with different text, and the researcher read "Start clicking!" Once the child clicked again, an image of a green field appeared, then nine training images, followed by the same green field and ending with the original cloud image with new text. When participants were distracted or overly slow and methodical in their clicking behavior on the training task (spending more than a couple of seconds per image), they were encouraged to "keep click, click, clicking to see the train go!" When children continued to click slowly or were inattentive to the event, the researcher had the child complete the training task again. Roughly 12% of the sample (26 participants) repeated the training.

Once the training block was over, the researcher said "Great job! Now you'll see more things on the screen, and you need to click the mouse to see them!" The researcher advanced to the next section with the space bar. Then participants clicked through the replication images, and

they saw the reach-over-barrier event with the ball, described above. The cloud photo appeared again, and the researcher said “Good job! Now you’re going to watch a video.” The researcher advanced the task with the space bar and participants watched the verbal pedagogy video. Then, the researcher said, “Now you’ll see more things on the screen, and you need to click the mouse to see them!” Participants then clicked through the main dwell time slideshow images (either seeing a press or wave event). The cloud image appeared at the end and the researcher said, “Here is the last slideshow!” Following this, participants clicked through the same dwell time images (i.e., either press or wave images) for a second time. At the end, the researcher said “Great job! You have completed the computer game.”

The child was asked to take off the headphones, and the researcher moved the computer to the other side of the table, so it was out of the way. Then the researcher said “I have a toy with me, and you can play with it for as long as you want. Let me know when you’re done playing with it.” The researcher reached behind their chair to a black zip bag containing the play object. They opened the bag and set it down in front of the child in the same orientation that it sat in front of the actor in the pedagogy videos and dwell-time slideshow. Children participated in the play session and then responded to comprehension questions as described in Study 1. Finally, children were directed to an area nearby where they were introduced to the research assistant who was leading the Head-Toes-Knees-Shoulders (HTKS) Executive Function task. Children were told “This is [name], and they will tell you about the next game!” They stood in front of the seated research assistant who led them through the HTKS task. The task ended when children either finished this task, or when they were no longer interested in participating. In all cases, they were offered a sticker and thanked for their participation.

3.6 Coding

3.6.1 Dwell-Time Exclusions

Research assistants tracked children's attentiveness during slideshow viewing by coding instances of looking away on a slide-by-slide basis. We used side-view video recordings to assess this, where there was both a view of the participant's line-of-regard and a view of the computer screen. The computer was tilted slightly toward the camera for visibility, and the camera recorded from a view at the opposite end of the table. Videos captured the child's upper torso, face profile, and hand on the mouse along with a view of the computer screen. For each participant, we coded individual slides as a '0,' '1,' or '2.' A score of '0' indicated that the child was looking away for the duration of the slide. A score of '1' meant that there was a portion of time when the child was looking at the slide and also a portion of time when they were looking away. A '2' indicated that the child looked continuously at the slide. We recorded look-aways for all slideshows that children viewed (i.e., training, replication, and main task) in the same manner. Coders did not know which pedagogical context condition (i.e. knowledgeable vs. naïve) children were participating when making look-away judgments. However, coders had visual access to the slides children were viewing, and therefore they were not blind to the expected action condition that children were participating in. Although we coded look-aways for the training and barrier-reach slideshows, I did not make exclusions on the basis of looking away for the analyses reported in the '3.10.1 Replication of Lab Sample' section below.[1] Among the two coders, Krippendorf's alpha value indicated high inter-rater reliability (0.91) for coders'

[1] Since researchers did not make this type of exclusion in the original study using the barrier-reach slideshow (Kosie & Baldwin, 2021), I did not make exclusions on this basis in the replication attempt.

look-away judgments. The dwell time look-away coding was meant primarily for the two main stimulus slideshows (depicting a press or a wave). If a participant had four or more '0's' recorded for a given slideshow (i.e., if they were looking away for >= 20% of the event), I took note of that participant for possible exclusion of data from analyses. For the first viewing of the slideshow, there were seven participants who looked away for four or more slides. For the second viewing, there were thirteen participants who looked away for four or more slides. I did not automatically exclude these participants but instead performed comparative analyses both with and without these participants for the main findings reported in the results (i.e. boundary effects, age effects, and effects of condition on dwell-time patterns). It turned out that excluding these participants made no difference in the significance or direction of any of the findings, and thus I retained their data in the results recorded below. There were an additional three participants who had 4+ look-aways as well, but their data were excluded according to the outlier cutoff process described in the 'data preparation' section below, rather than based upon the look-away criteria. One participant was looking away for the entirety of the pedagogy video, and their dwell-time data also contained outliers as described in the data preparation section below, so their data was excluded. There were ten participants who showed a partial look-away for the pedagogy video; their data were retained given that the key manipulation element included auditory cues (which all children heard through the headphones).

3.6.2 Additional Coding

In the analyses that included play outcomes and the main executive function measure (HTKS), the data were identical to and reflected the coding schemes described previously in Chapter 2 for Study 1.

3.7 Boundary Judgments

Each slide was classified as a boundary or non-boundary for the pressing and waving event slideshows. Two expert judges separately recorded certain slides as boundaries (such as when the right hand first left the side of the object, when the light turned on and off, when the trajectory of motion first changed downward for the press or sideways for the wave). There was general agreement regarding these judgments, and boundaries were set according to both event-related features (when the light turned on and off) and action-related features (when the hand left and returned to the side of the object or changed trajectory). Though experts made the boundary judgments here, data collection is underway to gain insight into naïve adult-participant judgments for our stimuli. I also note that boundary judgments made by experts tend to align with those made by naïve viewers in previous dwell-time research (Kosie & Baldwin, 2019; 2021).

Experts also made judgments regarding the hierarchical level depicted in boundary slides. Coarse-grain boundaries correspond to slides depicting the transition from one higher-level task to the next (such as the slide when pressing vs. waving culminates in the light-turning on), whereas fine-grain boundaries correspond to slides depicting the transition from one smaller-scale action to the next (such as the slide where the finger changes trajectory to begin the waving motion) (Newtson, 1973).

3.8 Accounting for Pixel Change

As in prior dwell time studies (Hard et al., 2011; Loucks & Baldwin, 2009; Kosie & Baldwin, 2019; Kosie & Baldwin, 2021), I accounted for slide-to-slide pixel change in an effort to determine if participants' dwell-time patterns were robust to surface differences in visual stimuli. That is, it was important to ensure that participants' dwell-time patterns were

representative of their responses to the content of the events that they viewed, not merely representative of their responses to motion characteristics, as indexed by pixel change from each slide to the subsequent adjacent slides. If children's dwell time patterns were robust to pixel change, then I expected to observe boundary effects even when controlling for pixel change. Following prior research (Loucks & Baldwin, 2009; Kosie & Baldwin, 2019), to create a pixel change index for the slideshows, I first ran each slideshow images through a convolution filter -- the Canny Edge Detector (Canny, 1986) -- prior to calculating pixel changes. This process, which is commonly used for detecting edges in images, allowed us to examine where the greatest motion-related visual changes in the event were in relation to the subject and object involved, rather than focusing simply on raw luminance changes. I downloaded two python files from: https://github.com/FienSoP/canny_edge_detector called 'canny_edge_detector.py' and 'canny_filter_and_save.py' (for generating and saving the filtered images). I used VS Code to run these files in order to generate our convolution-filtered images. Once these images were created, I calculated pixel change in R based upon the following algorithm:

$$\sum_{i=1}^{h}\sum_{j=1}^{w}\sqrt{\left(R_{Cij}-R_{Sij}\right)^2+\left(G_{Cij}-G_{Sij}\right)^2+\left(B_{Cij}-B_{Sij}\right)^2}$$

In this equation, the *R, G*, and *B* values represent red, green, and blue pixel colors. The *i* and *j* symbols characterize the coordinate values of a pixel within an image, while *h* and *w* represent the height and width of the image in pixels (Loucks & Baldwin, 2009). The *C* and *S* values track the change frame (current image) from the standard frame (previous image). Accordingly, although there were 21 slide images per slideshow, there were only 20 pixel change values, indicating the change between slides 1-2, 2-3, and so on. The convolution-filtered images and R file that I used to calculate pixel changes can be found in the supplementary materials for

the dissertation: https://osf.io/h6wat/files/osfstorage. A one-way omnibus ANOVA revealed that there were no significant differences in pixel change across the slideshows by expected action condition, $F(1, 36) = 1.51$, $\eta p^2 = .04$, 90% CI [0.00, 0.18] $p = 0.228$. (see Figure 3.1), indicating that motion parameters across the consistent and violating slideshows were generally equivalent despite the small essential differences associated with depicting a press versus wave action.

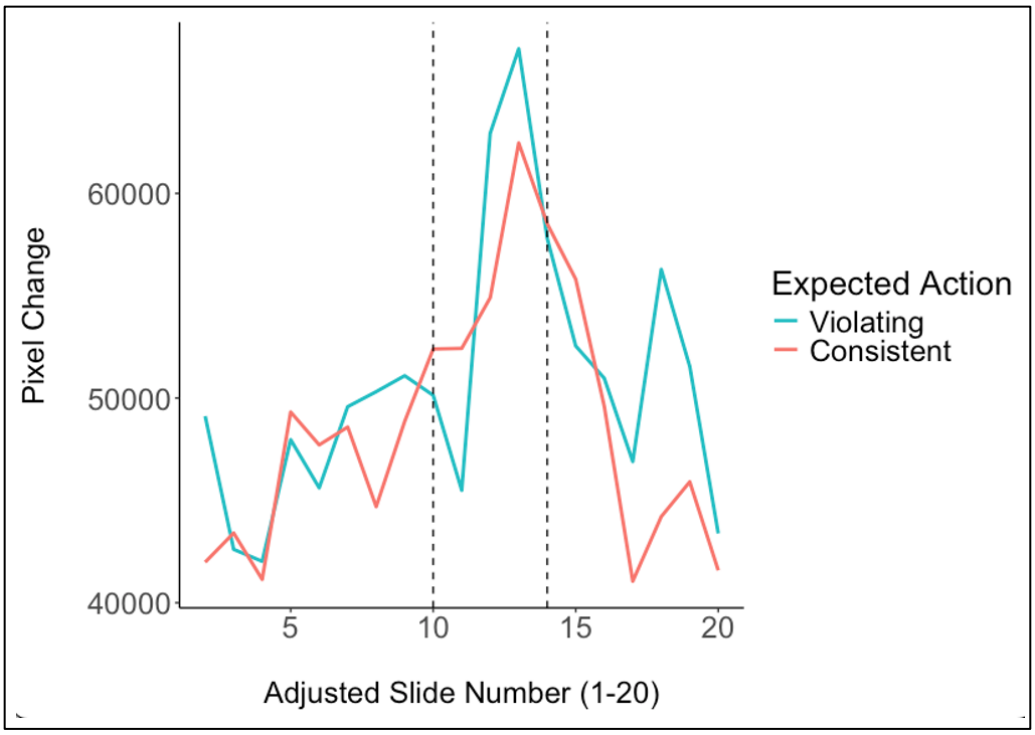


Figure 3.1. Slide-to-slide pixel change by expected action condition. The vertical dotted lines indicate the slides where the light first turns on (slide 10) and off (slide 14). Note that the y-axis values represent the pixel change between two slides, and the x-axis value tracks the first slide in the sequence (for example, slide 20 on the x-axis represents the pixel change between slides 20 and 21 from the filtered slide images).Because pixel change represents a single score per slide, error bars are not included.

3.9 Data Preparation

I began by removing dwell time data for the first and last slides for each event, as is customary for dwell time data given that these slides tend to represent atypical viewing (Hard et al., 2011). For each participant, if greater than 10% of their raw dwell times (recorded in seconds) were three or more standard deviations above the group mean for a given slideshow, they were excluded from any analyses which included that slideshow. Following these

exclusions, all other values that were three or more standard deviations above the mean of the group were Winsorized, meaning that they were given the mean-plus-or-minus-three-SD value. Because dwell times tend to be positively skewed (like looking time data more generally), a log10 transformation was applied to normalize the distribution of dwell-time scores following the exclusion and Winsorizing process. For the first viewing data, 12 participants were excluded via this process, bringing the total N to 208. For the second viewing data, 17 participants were excluded, leaving N = 203. In analyses comparing or averaging dwell-time scores across viewings, I maintained these exclusions.[2]

Furthermore, individual children's dwell-time patterns across each viewing of the slideshow were analyzed for the presence of a power-curve (i.e., a significant reduction in the individual participant's dwell-times as slides unfolded across the slideshow). Significant power curves are common but not universal in dwell-time data. I set a decision criteria that if at least 50% of children displayed a significant power curve, then residuals of the log-transformed dwell times from the power curve would be computed for each participant for every slide in each of the slideshows. Twenty four percent of participants showed a significant power curve for the first viewing and 19% showed a significant power curve for the second viewing (20% showed a significant power curve when dwell times were averaged across viewings). As this did not meet

---

[2]Following the exclusion process as described in-text, I received feedback that the typical order for handling exclusions when working with skewed data is to address skewness prior to addressing outliers. Therefore, I conducted a secondary exclusion process wherein the dwell time data were first log transformed, and then exclusions were made and additional outliers Winsorized as described in the text. This process indicated that only 3 participants should be excluded for the first viewing data (as opposed to 12 reported above), and only 9 participants should be excluded for the second viewing data (as opposed to 17 reported above). These were not new exclusions, in that the outlier process reported in-text included these participants already; however, I opted to re-analyze the data following this updated exclusion process (with 12 participants excluded total when dwell times were averaged across two viewings). Relevant results remained consistent with the dissertation reporting both for preliminary analyses, and for reporting related to research questions 1-4 in Study 2.

our pre-set threshold for creating residuals, I decided not to use residuals when examining related outcomes (i.e., repeated measures such as changes between the first and second viewing of the event). Thus, the analyses reported in the results below used only log-transformed dwell times (rather than using residuals of log-transformed dwell times for repeated measures). This decision reflects: a) that the sample failed to meet the threshold for significant power curves required for creating residuals, b) that residualization removes considerable between-subject variance and most of the predictions concerned between-subject conditions (which cannot be appropriately tested with residualized dwell times), and c) that a precautionary set of analyses revealed that using log-transformed outcomes did not differ from using residualized outcomes in relevant cases regardless.

Results

The approach to analyses for Study 2 was generally the same as that reported at the start of the results section for Study 1.

3.10 Preliminary Analyses

3.10.1 Replication of Lab Sample: Reach-Over-Barrier Event

Following the training task in our procedure, but before watching the verbal pedagogy video, all participants viewed a secondary event depicting a person reaching over a barrier to grab a ball. This event was used in a replication attempt as follows. In Kosie and Baldwin (2021), researchers used an identical event in their in-lab procedure with 31 participants (M = 2.77 years; range 2.5-4.5 years). When I accessed it, the Kosie and Baldwin data already accounted for outlier exclusion measures. I also note that the dwell time data in the Kosie and Baldwin sample was averaged across two viewings of the reach-over-barrier event whereas our data represented only the first viewing of the event (because children in the dissertation sample

only viewed this event once). As a result, there is reason to expect some systematic – although potentially minor – dwell-time differences across studies, given prior evidence of predictable reorganization in dwell times across repeated viewings of the same event. Lastly, prior to beginning analyses, I subset our sample to most closely match the age range reflected in their study.

Following age sub-setting and outlier exclusions, the resulting replication sample contained data from 71 participants (M = 3.31; range = 2.8-4.5 years). Using only these data, I fit a linear mixed-effects model to assess the extent to which participants dwelled longer at boundary vs. non-boundary slides. This is a robust finding from prior dwell-time research (Hard et al., 2011; Kosie & Baldwin, 2019; Meyer et al., 2011; Ross & Baldwin, 2015); enhanced dwelling to boundary slides appears to indicate that participants are sensitive to the unfolding structure of events. The model predicted dwell time from slide type (boundary vs. non-boundary) with participant age and their interactions as fixed effects with random intercepts for subjects and random slopes for slide type. Log10 dwell times were indeed significantly longer to boundary slides (M = 2.85, SD= 0.31) over non-boundary slides (M = 2.78, SD = 0.27) for the present sample, β = -0.07, SE = 0.02, $t(263)$ = -3.82, $p < .001$, as in the original Kosie & Baldwin, 2021 sample. Anticipated age effects, where younger children viewed longer on average than older children in the Kosie and Baldwin (2021) sample, failed to reach significance in the present sample, β = -0.01, SE = 0.01, $t(73)$ = -1.77, p = .081. It is worth noting that analyses of age effects reported later with dwell-times from our full sample viewing the expected action slideshows did display statistically significant replication of the Kosie and Baldwin (2021) age effect. Thus, the present data demonstrated the anticipated pattern of viewing (boundary

advantage effect) as the prior in-lab sample for the same event with participants of a similar age range. See Figure 3.2 for a slide-by-slide dwell time comparison by sample.

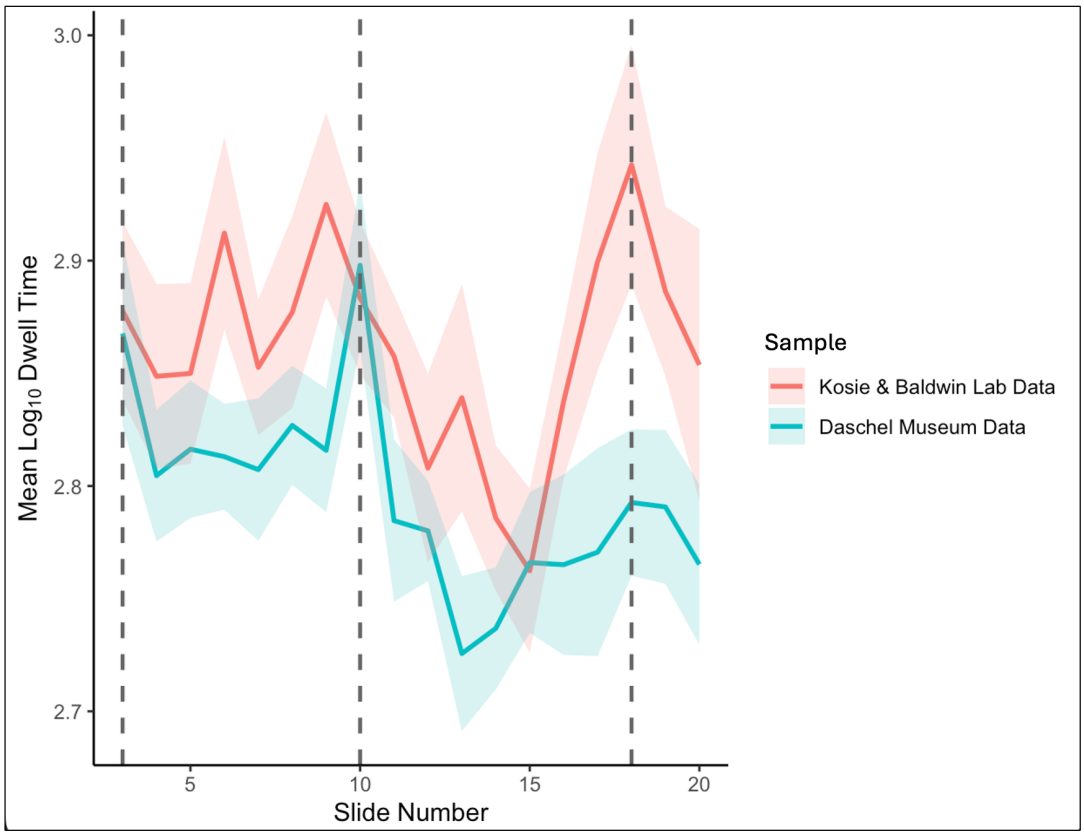


Figure 3.2. A comparison of dwell times by slide for the reach-over-barrier event including the Kosie & Baldwin (2021) lab sample and the Daschel museum sample. Vertical dashed lines indicate boundary slides, and shaded regions represent one standard error. Age range 30-53 months.

3.10.2 Replication of Age Effects: Expected Action Events

To investigate effects of age on dwell time patterns, I initially included age in one of the main confirmatory analyses as follows. A linear mixed-effects model was fit containing slide type (boundary versus non-boundary), centered pixel change, and age as fixed effects, plus their interactions, with random intercepts for subjects and random slopes for slide type. Age was associated with participants' overall log10 dwell times, with 3-to-4-year-olds (M = 2.76, SD =0.42) dwelling for longer on average than 5-to-7-year-olds (M = 2.65, SD = 0.22), β = -0.003, SE = 0.001, t(223.9) = -4.28, p < .001. However, age did not show a significant relationship to

the influence that slide type and pixel change had on children's dwell-time patterns. For this reason, I did not include age as a main variable in further analyses.

3.11 Confirmatory Analyses: Research Question 1

Moving on from data processing and preliminary analyses, I now turn to the main confirmatory analyses for dwell time patterns with the full sample. Confirmatory analyses probed boundary advantage differences across conditions (while controlling for pixel change). Exploratory analyses tested effects of pedagogical context and expected action condition on dwell time scores (including effects of these factors on dwell times related to first and second viewing). Exploratory analyses also examined relations between dwell time patterns and play behaviors, taking pedagogical context and actions expectations into account. For the following results (where relevant), *boundary* slide type, *knowledgeable* verbal pedagogy, and *consistent* expected action group are all treated as the reference group when fitting models.

3.11.1 ***Research Question 1:*** *To what degree will children's dwell time patterns replicate prior dwell time findings? That is, will analyses reveal a significant boundary advantage, as predicted (i.e., significantly higher dwell times during boundary regions relative to non-boundary regions) when controlling for pixel change?*

For analyses related to research questions 1, 2, and 4, dwell time scores were averaged across first and second viewing for each participant. I also ran analyses separately by first and second viewing, and the pattern of results remained consistent regardless of viewing. A linear mixed-effects model was fit containing slide type (boundary versus non-boundary) and centered pixel change, with random intercepts for subjects and random slopes for slide type. As predicted and replicating prior dwell-time research, log-transformed dwell times were significantly higher

to boundary *(M* = 2.72, *SD* = 0.22) relative to non-boundary slides *(M* = 2.69, *SD* = 0.21), β = -0.04, *SE* = 0.01, *t*(3163) = -7.24, *p* < .001 (Table 3.1).

Table 3.1. *Linear mixed-effects model output, with slide type (boundary vs. non-boundary) predicting log10 dwell times while controlling for pixel change.*

	Estimate	*SE*	*df*	*t*	*95% CI Lower*	*95% CI Upper*	*p*
Intercept	2.73	0.01	196.33	231.59	2.70	2.75	< .001
Slide Type	-0.04	0.01	3162.73	-7.23	-0.05	-0.03	< .001
Pixel Change	-8.61	1.38	3474.61	-6.23	-11.32	-5.90	< .001

Pixel change was significantly negatively related to dwell time, β = -8.61, *SE* = 1.38, *t*(3475) = -6.23, *p* < .001, which was somewhat unexpected, as one might expect greater pixel change to be positively associated with heightened dwell time scores. (Table 3.1 and Figure 3.3). For example, a higher concentration of pixels in a stimulus has been linked to higher salience, which is predictive of increases in reflexive attention to a given region (Kooiker et al., 2016; Henderson, 2003), and increases in pixel change from slide-to-slide indicate increased motion change, which tends also to be positively associated with longer dwelling (Hard, et al., 2011). Nonetheless, effects of slide type (i.e., enhanced dwelling at boundaries relative to non-boundaries) were robust even when controlling for pixel change, and slide type did not interact with pixel change in a model containing the same structure as the one reported above besides testing this interaction, *p* = .444. In any case, these findings clarify that dwell time patterns in the present study were robust even when controlling for the relationship between pixel change and dwell time scores.

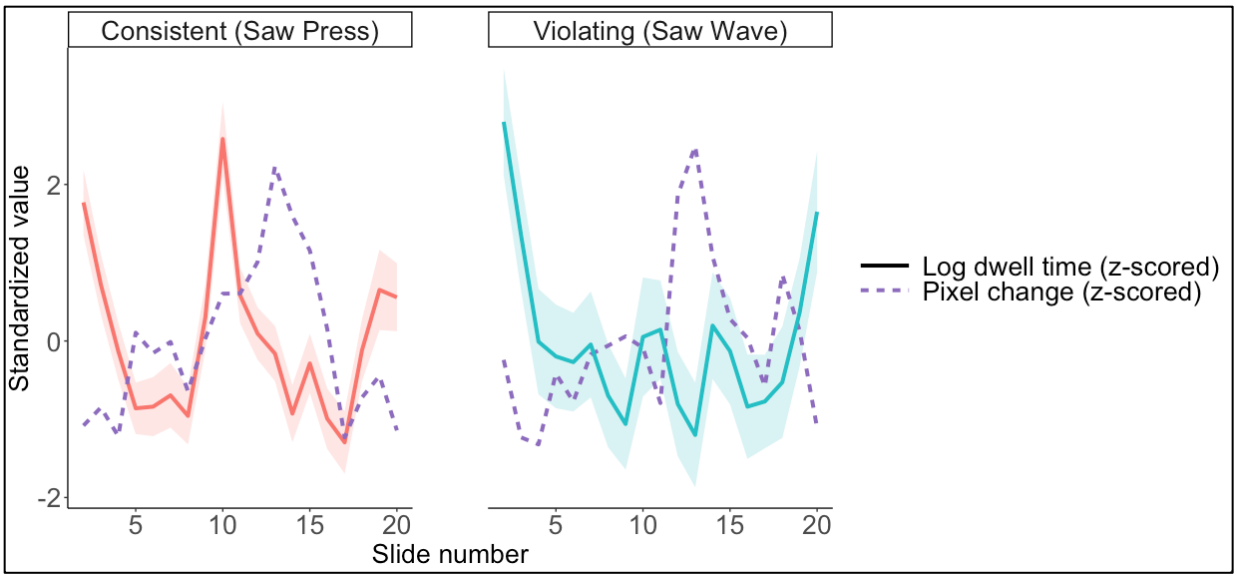


Figure 3.3. A comparison of pixel change and log10 dwell time patterns (with standard error ribbons) by expected action group.

Additionally, the initial boundary versus non-boundary dwell time difference was driven by the *consistent expected action* group. When looking just at the data from this group using the same model structure as above, the pattern of results was the same; dwell times were significantly longer for boundary versus non-boundary slides, $\beta =$ -0.06, $SE = 0.01$, $t(1593) =$ -7.12, $p < .001$. When looking only at the *violating expected action* group, however, an unexpected trend emerged wherein a boundary advantage was not observed, $p = .608$ (Figure 3.4).

As noted in the boundary judgments section above (Section 3.7), both coarse and fine-grained boundaries were represented in the events depicted in the expected-action slideshows, differing in the extent to which larger (coarse-grained) or smaller (fine-grained) event transitions were depicted by a given slide. The well-documented boundary advantage reflects a tendency for viewers to display longer dwelling on boundary versus non-boundary slides, whereas the hierarchical advantage reflects a tendency for viewers to display longer dwelling on boundary slides depicting coarse-grained relative to fine-grained-level boundary content. The hierarchical

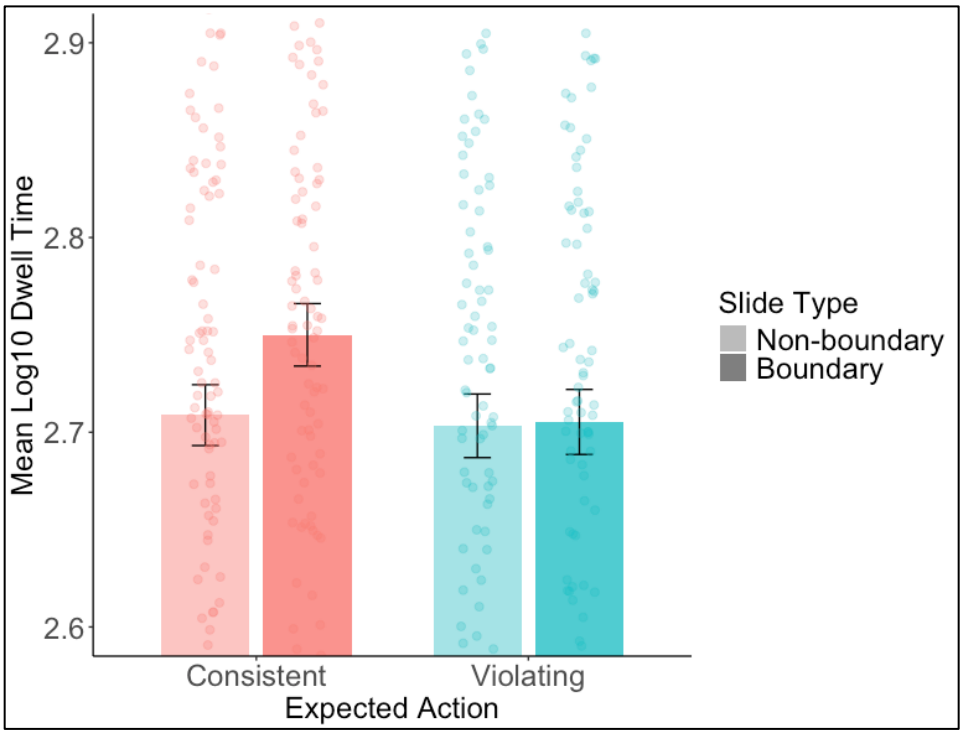


Figure 3.4. Mean log10 dwell time by expected action group (consistent vs. violating) and slide type (boundary vs. non-boundary) with standard error bars.

advantage is common, though not universal, in dwell-time research. Therefore, I also undertook analyses related to the hierarchical advantage (see Analysis Supplement document for Chapter 3 in the supplementary materials: https://osf.io/h6wat/files/osfstorage). This pattern of findings remained consistent with the overall boundary advantage findings reported above; that is, a hierarchical advantage was observed for the *consistent* but not *violating expected action* group. Because the hierarchical advantage findings paralleled the (typically more robust) general boundary advantage findings, I opted not to include the hierarchical boundary variable in further analyses.

3.11.2 Research Question 1 Summary

On the whole, findings related to research question 1 indicated that—controlling for pixel change—children in the *consistent expected action* group systematically deployed enhanced attention to boundary over non-boundary slides, as expected. In contrast, children in the *violating*

expected action group failed to display the predicted boundary advantage in their dwell-time scores, which was possibly related to the expectation-violation they were experiencing, rendering it difficult for them to detect event boundaries.

3.12 Exploratory Analyses: Research Questions 2-3

3.12.1 ***Research Question 2:*** *To what degree will children's a priori expectations regarding object-related action influence their dwell-time patterns, both directly and in interaction with the verbal pedagogy with which the object is introduced (i.e., knowledgeable vs. naïve conditions)?*

To assess the joint impact of verbal pedagogy and action expectations on children's dwell times, a linear mixed-effects model was fit containing slide type (boundary versus non-boundary), verbal pedagogy group (knowledgeable versus naïve), and expected action group (consistent versus violating) as fixed effects plus all of their interactions, with random intercepts for subjects and random slopes for slide type. As before, log transformed dwell times were significantly higher to boundary over non-boundary slides ($p < .001$). While neither verbal pedagogy nor expected action group presented fixed effects on their own, expected action interacted with slide type (Table 3.2). As discussed in relation to research question 1, this interaction between slide type and expected action group emerged because a boundary advantage effect occurred in the *consistent expected action* group but not the *violating expected action* group (see Figure 3.4 above and Table 3.3). Related to verbal pedagogy, boundary advantage effects did not differ by verbal pedagogy group (i.e., both knowledgeable and naïve verbal pedagogy groups displayed a significant boundary advantage) (Table 3.4 and Figure 3.5).

Table 3.2. *Linear mixed-effects model output, with slide type, verbal pedagogy group, and expected action group predicting log10 dwell times.*

	Estimate	*SE*	*df*	*t*	95% *CI Lower*	95% *CI Upper*	*p*
Intercept	2.71	0.02	190.45	113.81	2.67	2.76	< .001
Slide Type	-0.04	0.01	3011.87	-3.75	-0.06	-0.02	< .001
Verbal Pedagogy	0.03	0.03	190.48	1.05	-0.03	0.10	.295
Expected Action	-0.01	0.03	192.35	-0.18	-0.07	0.06	.855
Slide Type × Verbal Pedagogy	-0.02	0.01	3012.66	-1.13	-0.04	0.01	.257
Slide Type × Expected Action	0.04	0.01	3011.87	2.74	0.01	0.07	.006
Verbal Pedagogy × Expected Action	-0.03	0.05	192.39	-0.69	-0.12	0.06	.490
Slide Type × Verbal Pedagogy × Expected Action	0.01	0.02	3012.26	0.38	-0.03	0.05	.706

Table 3.3. *Bonferroni-corrected contrasts showing the difference in boundary versus non-boundary slides across expected action groups, highlighting their* ***two-way interaction****.*

Expected Action	**Slide Type Contrast**	**Estimate**	*SE*	*df*	*t*	*p (Bonf.)*
Consistent	boundary vs. nonboundary	0.05	0.01	3012.66	6.57	< .001
Violating	boundary vs. nonboundary	0.00	0.01	3011.87	0.42	.676

Table 3.4. *Bonferroni-corrected contrasts showing the difference in boundary versus non-boundary slides across verbal pedagogy groups. Note that the two-way interaction between slide type and verbal pedagogy was not significant in the reported findings (p = .257; Table 3.2).*

Verbal Pedagogy	**Slide Type Contrast**	**Estimate**	*SE*	*df*	*t*	*p (Bonf.)*
Knowledgeable	boundary vs. nonboundary	0.02	0.01	3011.87	2.57	.010
Naive	boundary vs. nonboundary	0.03	0.01	3012.69	4.43	< .001

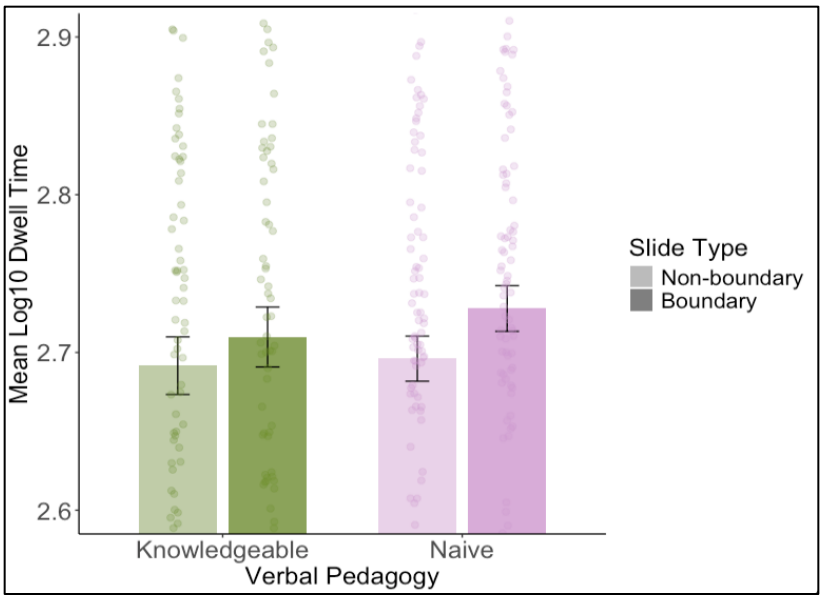


Figure 3.5. Mean log10 dwell time by verbal pedagogy group (knowledgeable vs. naive) and slide type (boundary vs. non-boundary) with standard error bars.

3.12.3.2 Research Question 2 Summary

Children in the *consistent expected action* group, as opposed to those in the *violating expected action* group, showed significantly longer dwell times to boundary over non-boundary slides (i.e., a stronger boundary advantage). Such a boundary advantage did not differ by verbal pedagogy group. All in all, results related to research question 2 demonstrated that children's dwell time patterns were directed especially by the characteristics of the event itself (i.e., consistent vs violating).

3.12.3 *Research Question 3:* *To what extent will children's dwell-time patterns change from first to second viewing, and to what degree will such changes across viewings interact with pedagogical context (i.e., knowledgeable vs. naïve conditions) and action expectations (i.e., consistent versus violated expectations)?*

As explained in the method, children viewed the same expected action slideshow twice. Viewing (first versus second) was included as a dwell time measure in part because prior dwell time findings have indicated that attention reorganizes upon multiple viewings of the same event

(Kosie & Baldwin, 2019). This reorganization tends to be especially pronounced when an event contains novel elements or when a participant is unfamiliar with an event (Hard et al., 2019). Therefore, I included the following exploratory approach related to research question 3 on the basis of prior dwell time findings indicating viewing as an influence on the degree of attention reorganization that is expected to occur in tandem with event properties.

To assess the joint impact of verbal pedagogy, action expectations, and viewing (first versus second) on children's dwell times, a linear mixed-effects model was fit containing slide type, verbal pedagogy group, expected action group, and viewing as fixed effects plus all of their interactions, with random intercepts for subjects and random slopes for slide type. I discussed several of the included variables and their interactions already in research questions 1 and 2, and the pattern of findings remained consistent. For this reason, I focus here on effects of viewing. While viewing was not a fixed effect on its own, a significant interaction between viewing and verbal pedagogy emerged, $\beta = -0.04$, $SE = 0.02$, $t(7561) = -1.98$, $p < .05$ (Table 3.5). Those in the *naïve verbal pedagogy* group showed a significantly greater difference between first versus second viewing ($p < .001$) when compared with those in the *knowledgeable verbal pedagogy* group ($p = .411$) (Table 3.6 and Figure 3.6).

Table 3.5. *Linear mixed-effects model output, with slide type, verbal pedagogy group, expected action, and viewing (first versus second) predicting log10 dwell times.*

	Estimate	*SE*	*df*	*t*	95% *CI Lower*	95% *CI Upper*	*p*
Intercept	2.73	0.02	266.41	112.20	2.68	2.78	< .001
Slide Type	-0.04	0.01	6951.30	-2.80	-0.06	-0.01	.005
Verbal Pedagogy	0.03	0.03	255.81	0.78	-0.04	0.09	.439
Expected Action	-0.03	0.03	289.15	-1.03	-0.10	0.03	.305
Viewing	-0.01	0.01	7555.42	-0.94	-0.04	0.01	.347
Slide Type × Verbal Pedagogy	-0.02	0.02	6905.38	-0.87	-0.05	0.02	.385
Slide Type × Expected Action	0.04	0.02	6991.00	2.14	0.00	0.08	.032
Verbal Pedagogy × Expected Action	-0.02	0.05	267.36	-0.37	-0.11	0.08	.714
Slide Type × Viewing	0.02	0.02	7546.39	1.05	-0.02	0.06	.296
Verbal Pedagogy × Viewing	-0.04	0.02	7561.08	-1.98	-0.07	-0.00	.047
Expected Action × Viewing	-0.00	0.02	7553.20	-0.04	-0.04	0.04	.970
Slide Type × Verbal Pedagogy × Expected Action	0.02	0.03	6920.64	0.64	-0.04	0.07	.525
Slide Type × Verbal Pedagogy × Viewing	-0.01	0.03	7545.63	-0.20	-0.06	0.05	.840
Slide Type × Expected Action × Viewing	-0.01	0.03	7545.59	-0.40	-0.06	0.04	.686
Verbal Pedagogy × Expected Action × Viewing	0.03	0.03	7560.67	1.00	-0.03	0.08	.320
Slide Type × Verbal Pedagogy × Expected Action × Viewing	-0.03	0.04	7545.22	-0.86	-0.11	0.04	.393

Table 3.6. *Bonferroni-corrected contrasts showing the difference in log10 dwell time by viewing across verbal pedagogy groups, highlighting their* ***two-way interaction****.*

Verbal Pedagogy	**Viewing Contrast**	**Estimate**	*SE*	*df*	*t*	*p (Bonf.)*
Knowledgeable	First vs. Second	0.01	0.01	7643.63	0.82	.411
Naive	First vs. Second	0.04	0.01	7591.03	5.97	< .001

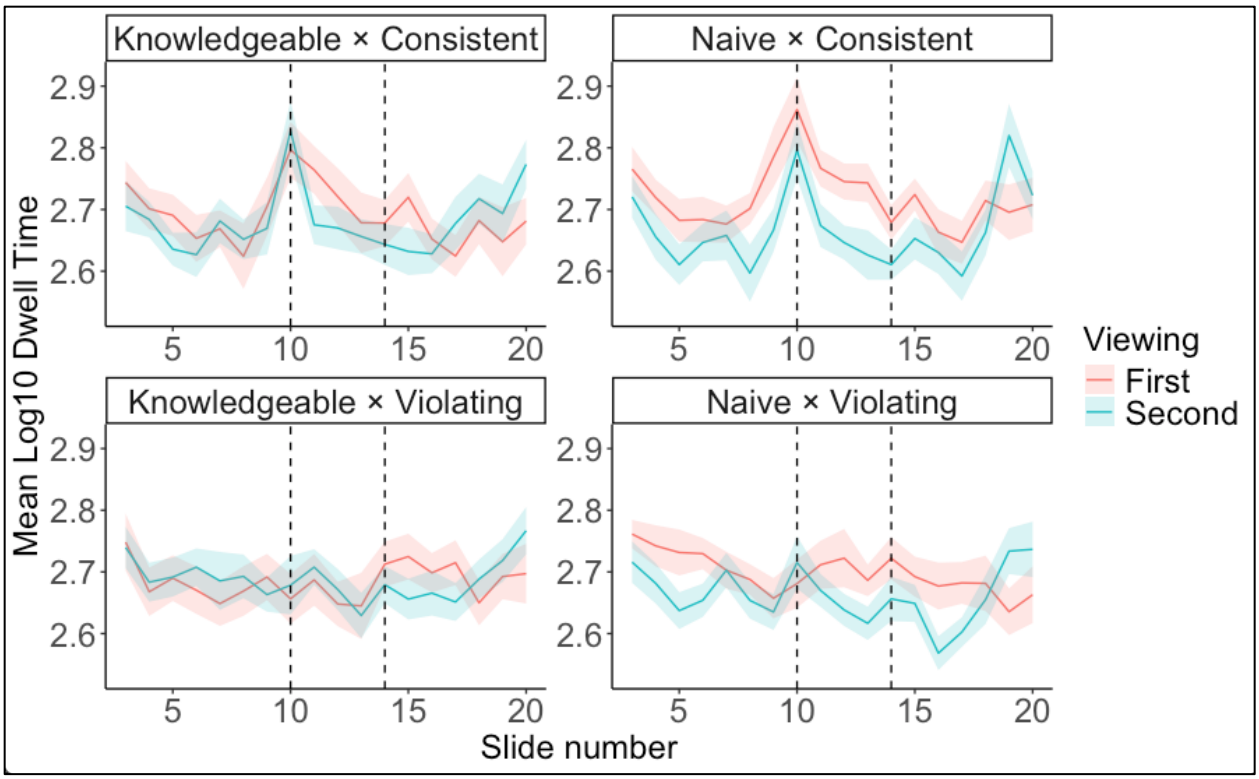


Figure 3.6. A comparison of mean-per-slide log10 dwell times by condition and viewing. The vertical dotted lines represent the time in the slides when the light first turned on (slide 10) and off (slide 14). The shaded ribbons represent one standard error from the mean.

3.12.4 Research Question 3 Summary

Changes in viewing were not related to the nature of the event that children saw (consistent versus violating); however, changes in viewing emerged in relation to the verbal pedagogy children heard. Those in the *naïve verbal pedagogy* group showed a significant difference in mean-per-slide log10 dwell times by viewing—with longer dwell times for the first versus the second viewing. Those in the *knowledgeable verbal pedagogy* group, on the other hand, did not show significantly different mean-per-slide log10 dwell times by first versus second viewing. These results demonstrate that hearing *naïve verbal pedagogy* (as opposed to hearing *knowledgeable verbal pedagogy*) led to a greater change in dwell times from first to second viewing.

3.12.5 ***Further (non a priori) Exploratory Research Question 4:*** *For those who saw the violating event and heard naïve verbal pedagogy, to what extent will increases in children's exploratory play behavior show relations to their dwell time patterns indicative of enhanced attention, and will such relations be impacted by executive function skills?*

For the following, I focused only on the *violating expected action x naïve verbal pedagogy* group, since this group showed the greatest degree of enhanced exploration in the results for Study 1. Within this group, I was interested in the extent to which EF would influence relations between dwell times and play behaviors. As a reminder, while EF did not by itself predict the proportion of time children spent with the demonstrated action or the number of unique actions they performed in Study 1, EF did interact with verbal pedagogy group and expected action group to influence both of these outcomes. Children higher in EF showed greater fidelity in their play (i.e., performing a greater number of unique actions in the *naïve x violating* group and performing a greater proportion of waves in the *knowledgeable x violating* group). EF also interacted with age to predict whether children in the violating expected action group performed any waves (with those both older and higher in EF being most likely to perform any waves). Thus, I focused on these three play outcomes (proportion of time with demonstrated action, number of unique actions, and any wave) in relation to EF to assess the extent to which they were together related to dwell-time patterns.

As reported in the preliminary analyses, age affected children's dwell times, with younger children exhibiting longer dwell times on average than older children. I first sought to understand if EF would be affiliated with dwell time changes when controlling for age within the *violating expected action x naïve verbal pedagogy group*. I fit a linear mixed-effects model with EF as a fixed effect and age as a covariate with random intercepts for subjects predicting

participants' log10 dwell times. However, EF was not significantly related to children's per-slide dwell times on its own when controlling for age β = -0.001, SE = 0.001, $t(44)$ = -1.27, p = .209

Next, I investigated the proportion of time with demonstrated action variable as follows: a linear mixed-effects model was fit containing slide type (boundary versus non-boundary), proportion of time with the demonstrated action, and EF (controlling for age) as fixed effects plus their interactions, with random intercepts for subjects and random slopes for slide type. No main or interaction effects emerged in this model, all ps > .075.

Similarly, I assessed the extent to which the number of unique actions variable was related to dwell time patterns in what was otherwise the same model as above besides including the unique actions variable rather than the proportion of time variable. While no fixed effects emerged for slide type, EF, or unique actions, there was a two-way interaction between unique actions and slide type, β = -0.06, SE = 0.02, $t(746)$ = -2.60, $p < .05$, as well as a three-way interaction between unique actions, slide type, and EF, β = 0.00, SE = 0.00, $t(746)$ = 2.60, $p < .05$ (Table 3.7). A follow up comparison—conditional slope contrasts--for the two-way interaction was not significant (p = .670) (Table 3.8), indicating that the number of unique actions on its own did not predict changes in boundary advantage patterns. However, the same type of follow up comparison for the three-way interaction revealed that for those lower in EF, demonstrating a boundary advantage was especially tied to performing a greater number of unique actions, while this was not the case for those higher in EF (Table 3.9 and Figure 3.7). That is, those lower in EF showed a greater boundary advantage increase as the number of unique actions they performed increased, while this was not true for those higher in EF.

Table 3.7. *Linear mixed-effects model output, with slide type, unique actions, and EF (with age as a covariate), predicting log10 dwell times.*

	Estimate	*SE*	*df*	*t*	95% *CI Lower*	95% *CI Upper*	*p*
Intercept	2.71	0.06	44.31	49.11	2.60	2.82	< .001
Slide Type	0.02	0.02	746.32	0.70	-0.03	0.06	.482
Unique Actions	0.07	0.05	43.25	1.53	-0.02	0.17	.133
EF (HTKS score)	-0.00	0.00	44.67	-0.25	-0.00	0.00	.807
Age	-0.01	0.00	43.00	-2.30	-0.01	-0.00	.027
Slide Type x Unique Actions	-0.06	0.02	746.32	-2.60	-0.10	-0.01	.010
Slide Type x EF (HTKS score)	-0.00	0.00	746.32	-0.92	-0.00	0.00	.360
Unique Actions x EF (HTKS score)	-0.00	0.00	43.35	-1.49	-0.00	0.00	.143
Slide Type x Unique Actions x EF (HTKS score)	0.00	0.00	746.33	2.60	0.00	0.00	.010

Table 3.8. *A contrast of slopes for slide type by unique actions, indicating that children's dwell time boundary advantage did not systematically increase or decrease as they performed a greater number of unique actions.*

Slide Type Slope Contrast	**Estimate**	*SE*	*df*	*t*	*p*
boundary vs. nonboundary	0.00	0.01	599.93	0.43	.670

Table 3.9. *A contrast of slopes for slide type by unique actions and EF level (-1SD, mean, +1SD), indicating that the boundary advantage increased as the number of unique actions increased for the low EF children, but not for others.*

EF Level	**Slide Type Slope Contrast**	**Estimate**	*SE*	*df*	*t*	*p*
−1 SD	boundary vs. nonboundary	0.04	0.02	746.33	2.50	.013
Mean	boundary vs. nonboundary	0.01	0.01	746.33	1.27	.205
+1 SD	boundary vs. nonboundary	-0.02	0.01	746.33	-1.35	.178

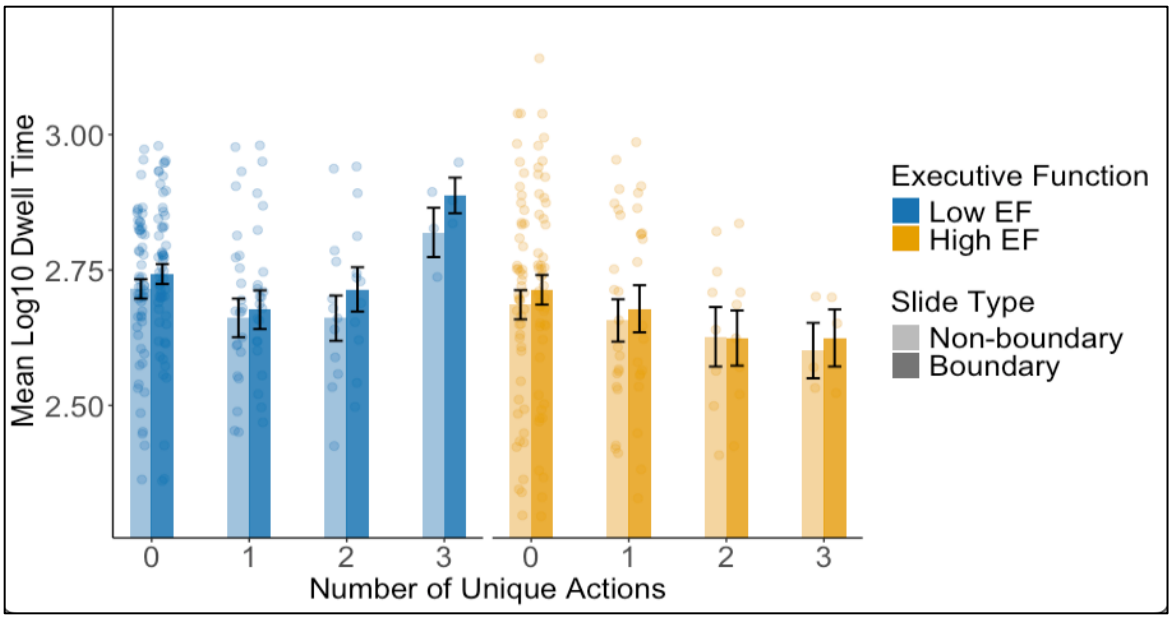


Figure 3.7. A comparison of log10 dwell times by unique actions, slide type, and EF group, with error bars (only data from the violating expected action x naïve verbal pedagogy group included). The HTKS scores range from 0-52, with the cutoff of 26 to distinguish participants who scored well on training and Part 1 of the task and moved on to Part 2 (designated High EF) from those who scored poorly on Part 1 or failed to complete it (designated Low EF).

Finally, I assessed relations between dwell time and the extent to which children performed any waves. A linear mixed-effects model was fit containing slide type (boundary versus non-boundary), whether any wave was performed, and EF (controlling for age) plus all of their interactions, with random intercepts for subjects and random slopes for slide type. While no other main or interaction effects emerged, there was a two-way interaction between EF and the performance of any waves, $\beta =$ -0.01, $SE = 0.00$, $t(44) = -2.65$, $p < .05$ (Table 3.10). A follow up comparison focusing on the impact of waving behavior by EF level in their relation to log10 dwell times revealed that for those high in EF, producing a wave was associated with shorter overall dwell times than not producing a wave (Table 3.11 and Figure 3.8).

Table 3.10. *Linear mixed-effects model output, with slide type, any waves, and EF (with age as a covariate), predicting log10 dwell times.*

	Estimate	*SE*	*df*	*t*	*95% CI Lower*	*95% CI Upper*	*p*
Intercept	2.72	0.05	44.36	56.79	2.62	2.81	< .001
Slide Type	-0.03	0.02	570.27	-1.27	-0.07	0.02	.205
Any Wave	0.11	0.07	42.97	1.48	-0.04	0.25	.147
EF (HTKS score)	0.00	0.00	43.36	0.96	-0.00	0.01	.342
Age	-0.00	0.00	43.01	-1.25	-0.01	0.00	.218
Slide Type x Any Wave	0.01	0.04	570.27	0.28	-0.06	0.08	.776
Slide Type x EF (HTKS score)	0.00	0.00	570.27	0.87	-0.00	0.00	.387
Any Wave x EF (HTKS score)	-0.01	0.00	43.55	-2.65	-0.01	-0.00	.011
Slide Type x Any Wave x EF (HTKS score)	-0.00	0.00	570.27	-0.37	-0.00	0.00	.713

Table 3.11. *Conditional contrasts indicating whether performing any waving influenced overall dwell times across EF levels.*

EF Level	**Wave Contrast**	**Estimate**	*SE*	*df*	*t*	*p*
−1 SD	Not Waving vs. Waving	-0.05	0.05	42.99	-1.00	.321
Mean	Not Waving vs. Waving	0.07	0.04	43.10	1.65	.106
+1 SD	Not Waving vs. Waving	0.20	0.07	43.11	2.90	.006

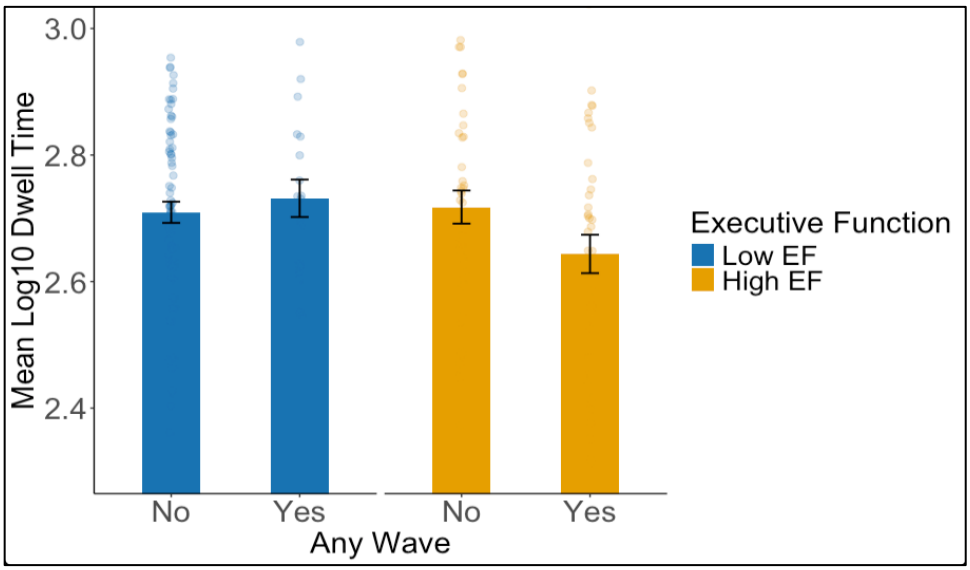


Figure 3.8. A comparison of log10 dwell times by whether any waves were performed and EF group, with error bars (only data from the violating expected action x naïve verbal pedagogy group included). The HTKS scores range from 0-52, with the cutoff of 26 to distinguish participants who scored well on training and Part 1 of the task and moved on to Part 2 (designated High EF) from those who scored poorly on Part 1 or failed to complete it (designated Low EF).

3.12.6 Research Question 4 Summary

For the *violating expected action x naïve verbal pedagogy* group, EF on its own was not related to children's log10 dwell times. The proportion of time children spent waving (either on its own or in relation to EF) was also unrelated to log10 dwell times. However, the number of unique actions related to dwell times via EF: children lower in EF who performed a greater number of unique actions also showed a greater increase in boundary advantage. Performing any waves also related to dwell times via EF; however, in this case, it was those higher in EF who showed differences in dwell times by waving behavior (those high in EF who waved showed shorter dwell times than those high in EF who did not wave). These results demonstrated that exploratory play behavior and executive function jointly influenced children's *pattern* of dwell times (in the case of unique actions) as well as children's *overall* dwell times (in the case of performing any waves). Enhanced attention to dwell time patterns was associated with greater exploratory behavior for certain children—as performing a greater number of unique actions especially mattered for lower EF children in the demonstration of a boundary advantage. In another case, performing any waves especially mattered for higher EF children in more quickly processing the overall details of the event.

Discussion

In Study 2, boundary advantage effects (wherein participants dwell longer on slides depicting boundary content of an event, as important causal features are taking place) were robust to pixel change. This means that children were patterning their attention during event viewing not merely on superficial image characteristics, but upon the structure of the event itself (e.g. what is causing the light to turn on, how was the toy held, what are the hands doing, etc.).

Those in the *consistent expected action* group (who saw an expected pressing action) showed this boundary advantage effect while those in the *violating expected action* (who saw an unexpected wave) group did not. This indicates that those in the violating group may have had some trouble detecting the overall structure of the event. There were also dwell-time differences related to pedagogical context, where those in the *naïve verbal pedagogy* group showed a larger drop in dwell times from first to second viewing of the same event. Finally, there was a relation between exploratory behavior, executive function, and enhanced attention, indicating that performing unique actions related to an increased boundary advantage for children lower in EF, whereas performing any waves related to lower overall dwell times for children higher in EF.

Taken together, the evidence from Study 2 suggests that children's dwell times were differentiated in subtle ways, across verbal pedagogy, expected-action group, play behaviors, and EF. Children's exploratory play behavior was associated with enhanced attention—with increases in exploration (unique actions) especially mattering for whether lower EF children showed greater viewing to boundaries. On the other hand, performing the violating action at all (any wave) especially mattered for whether higher EF children showed quicker processing across slides. In Chapter 4, I turn to an in-depth look at a) a comprehensive summary of findings across both Studies 1 and 2, b) future implications, and c) limitations and concluding takeaways.

CHAPTER IV

GENERAL DISCUSSION

4.1 Overview

My overarching aim in the dissertation was to investigate possible joint influence of verbal pedagogy, action expectations, and executive skill on children's exploratory play (Study 1) and attentional patterns (Study 2). Below, I summarize main findings, discuss limitations, and address implications and future directions for this work.

4.2 Summary of Key Findings

4.2.1 Study 1 Exploratory Play: Proportion of Time with Demonstrated Action

Prior "double-edged sword of pedagogy" findings (e.g., Bonawitz et al., 2011) led me to predict that knowledgeable pedagogy ("This is how my toy works") would induce children to devote a greater proportion of their time to enacting the action they'd seen demonstrated, relative to naïve pedagogy ("I wonder how this toy works"). My findings failed to confirm this prediction, likely largely because the novel object I utilized in the dissertation (the light) powerfully afforded one particular action (pressing), leading children across all conditions to devote considerable time to the pressing action, even when it wasn't demonstrated. As well, the proportion of time spent pressing reached near-ceiling levels when averaging across knowledgeable pedagogy conditions, meaning that a ceiling effect may have undercut replication of the Bonawitz, et al. predicted increase in proportion-demonstrated action for knowledgeable versus naïve conditions. In contrast to the pedagogical context variable, the expected action variable did strongly influence children's proportional play, as expected. With just a handful of exceptions, nearly all children's play was devoted almost exclusively to pressing when pressing was what they'd witnessed (consistent expected action group), whereas witnessing a violation of the pressing expectation (waving) significantly reduced the proportion of time children spent pressing. These findings clarify that, as

expected, action expectations can powerfully constrain children's exploration, and via this mechanism, reduce their likelihood of discovering something new (such as learning that a wave would also activate the light).

4.2.2 Study 1 Exploratory Play: Number of Unique Actions

The prior "double-edged sword of pedagogy" findings (Bonawitz, et al., 2011) also motivated the prediction that naïve pedagogy would elicit enhanced exploratory behavior, as indexed by more unique actions (beyond presses and waves) on the light, relative to knowledgeable pedagogy. My findings confirmed the prediction. At the same time, a significant interaction emerged between verbal pedagogy and action expectations on the number of unique actions: children showed the exploratory increase in unique actions only when naïve pedagogy and violation of action expectations were combined. Clearly, these sources of influence operated jointly to influence children's exploratory activities. It is also worth noting that the joint influence of naïve pedagogy and violation of expectations enhanced children's exploratory activities independent of age, suggesting stability of this pattern of joint influence on unique actions throughout the relatively broad age span included in the dissertation research.

4.2.3 Study 1 Exploratory Play: First Action

In an exploratory analysis, I examined the possibility that witnessing waving when pressing was expected might have induced children to subsequently press the light themselves out of an exploratory response; that is, by way of testing whether pressing might be effective in turning the light on, as well as waving. Indeed, for those in the *violating expected action* group, first pressing and later waving (as opposed to first pressing and never waving) was related to the number of unique actions that children performed. Pressing-plus-waving seemed thus to serve as an additional indicator of exploration among those in the *violating expected action* condition.

4.2.4 Study 1 Exploratory Play: Relations to Executive Function and Temperament

Children's executive skill was systematically related to the nature of their play. In particular, in the *violating expected action* group, higher EF was associated with a greater proportion of waves. In contrast for this same group, EF was not systematically associated with the number of unique actions children displayed. However, also in the *violating expected action* group, those higher in Extraversion/Surgency and Effortful Control devoted increased time to waving. Thus, strong EF skills and EF-related temperament dimensions seemed to enhance the fidelity of children's play in relation to the demonstration they'd witnessed: they showed more waving when waving was what they'd witnessed than those with weaker EF skills. This seemed to be especially true with increasing age, as each month increase in age conferred about a 5% increase in prioritizing waving with each point increase in EF. Lastly, exploratory analysis of the consistent expected action group revealed a finding of particular interest: when children had witnessed pressing (consistent with their expectation), higher EF was associated with children being more likely to inadvertently discover that waving (in addition to pressing) activated the light, suggesting that strong EF skills also enhanced children's detection of relevant new causal information as they were in the midst of active play.

4.2.5 Study 1 Summary

Together, Study 1 findings showcased that children's exploratory activities arose as a joint product of the pedagogical context, their *a priori* expectations about the object to be acted upon, the action evidence they actually observed, and the executive skills they brought to the learning situation.

4.2.6 Study 2 Dwell Time: Boundary Advantage

Based on prior dwell-time research (Hard, et al., 2011; Kosie & Baldwin, 2021), I expected children to display a boundary advantage in their dwell-time patterns across conditions; that is, I predicted longer dwelling to slides depicting causally-relevant boundary than non-boundary regions as they viewed the action demonstration underway. My findings confirmed this prediction, even when controlling for a pixel-change-based estimate of surface motion differences, clarifying that the boundary advantage was indeed indicative of children's analysis of the causal structure of the action they witnessed. Strikingly, however, a significant interaction between expected action group and slide type (boundary vs. non-boundary) revealed that the boundary advantage was in fact confined to the *consistent expected action* group (who saw a press). This finding suggested that many, or perhaps even all, children viewing the violating expected action (waving) may not yet have achieved an analysis of the causal structure of the unexpected waving action during their viewing of that action.

4.2.7 Study 2 Dwell Time: Viewing (First versus Second)

Prior dwell-time research documented that adults' processing of a novel motion stream undergoes rapid, systematic reorganization (Kosie & Baldwin, 2019). On first viewing, adults' dwell-time patterns have been shown to lack a boundary advantage -- suggesting that they aren't yet sensitive to the causal structure of the event – which strikingly often emerges already on second viewing. I predicted a comparable reorganization in children's dwell-time patterns on second versus first viewing in Study 2, and I suspected that this reorganization would be especially pronounced in the violating expected action group. Although children in the naïve verbal pedagogy group displayed an overall tendency toward shorter dwelling on second relative to first viewing, the specific pattern of reorganization I predicted did not emerge in children's attentional patterns

in any of the four experimental conditions. That is, children did not show systematic evidence of an increased boundary advantage on second versus first viewing in any pairing of verbal pedagogy and action expectations, but they did show evidence of an overall decrease in dwelling across first versus second viewing by verbal pedagogy group.

4.2.8 Study 2 Dwell Time: Relations to Exploratory Play in the Violating Expected Actions Group

Keeping in mind that sample sizes were small for these exploratory analyses, children's dwell-times in the violating expected action group displayed no systematic relationship to either the proportion of time they spent with the demonstrated action or children's EF scores alone. However, shorter overall dwell times were associated with performing any waves for high EF children, and a larger boundary advantage was associated with a greater number of unique actions performed by low EF children.

4.2.9 Study 2 Summary

Both pedagogical context and action expectations influenced children's dwell times. In particular, the *naïve verbal pedagogy* group and the *consistent expected action* group drove the largest boundary-related and viewing-related differences in dwell times, whereas those in the *naïve verbal pedagogy x violating expected action* group displayed systematic changes in dwell times in relation to both executive function and exploratory play behaviors.

4.2.10: Study 1 and Study 2 Integration

Taken together, the findings from both Study 1 and Study 2 point to the combination of naïve verbal pedagogy and seeing a violation of an expected action as primary drivers of exploratory play and attentional reorganization upon multiple viewings of an event. In fact, naïve verbal pedagogy related to attentional reorganization for both consistent and violating actions. This suggests that hearing a naïve verbal pedagogy statement primes children to update their attention

allocation and to explore more broadly when paired with a violating causal event. Moreover, EF skill and scoring higher on EF-related temperament dimensions were associated with exploratory activity displaying greater sensitivity to both quality of pedagogy and the evidence children encountered (i.e., consistent with, or in violation of, their expectations). Collectively, these findings point to EF skill conferring considerable advantage in enhancing children's knowledge-updating processes.

4.3 Strengths, Limitations, Remaining Questions

The methodological approach I took in Studies 1 and 2 reflected some noteworthy strengths: I collected data in a naturalistic museum setting rather than a lab setting, and I probed multiple aspects of children's functioning in the study, including their exploratory play activities, their attentional patterns, and their executive skill. The richness of this approach enabled me to assess the joint impact of different dimensions of influence on children's learning. My multi-dimensional approach also enhances the generalizability of the findings in that it captures something of the complexity children confront in their everyday learning. Nevertheless, the approach I took had limitations that will need to be addressed in future research. I now turn to discussing these issues.

4.3.1 Sample Size Limitations

As described in the Method for Studies 1 and 2, a power analysis revealed that to achieve effects with at least 0.9 power, the ideal sample was N = 231. At N = 223, our sample was not far short of that, nor was our degree of power (0.88); however, ideally, we would have achieved a full sample that accommodated the presence of exclusions while still retaining 0.9 power to reduce Type II error. While studies with children in both the dwell time and exploratory play literatures often have smaller samples, our two-and three-way interactions should be interpreted only within the scope of our sample. For example, 24 participants were excluded from the executive function

measure (HTKS), and exclusions on that scale were similar for the dwell time data as a result of outlier processing. This meant that for analyses involving these variables, between-group comparisons were closer to 50 participants per group rather than the ideal 60+ participants per group. All in all, my findings related to executive function and dwell time patterns should be regarded as exploratory and in need of replication.

4.3.2 Novel Object Limitations

My initial goal in the creation of the novel object utilized in the research (the light) was to craft a manipulation that was both controlled and subtle. That is, I aimed to ensure children were responding to a single visual affordance – in this case, a central light-up button – that could be acted upon in multiple ways to produce the same outcome. I initially considered including additional 'distractor' features so as not to artificially constrain children's exploratory focus to the central feature of the object (its ability to light up). For example, it is common within the larger literature on children's exploratory play to utilize visually complex objects that produce multiple interesting causal outcomes (Bonawitz et al. 2011; Gweon et al., 2014). I opted to include only one visually interesting feature (the button on the light) that could be acted upon in multiple ways in order to simplify analysis of children's response to sources of influence in relation to that one feature. For example, it would have been both difficult and beyond the main aims of the research to attempt to explain why children in the *violating* versus *consistent expected action* groups might display differences in play based upon distractor items. However, this decision likely had the downside of increasing children's near-ceiling level of focus on pressing the light across conditions except when the event demonstration introduced an action (waving) that violated children's pressing expectation. That is, the salience of the pressing action was likely amplified by the fact that it was the only action visually afforded by the light. It will be important to replicate the

dissertation findings with a greater variety of objects that display a greater range of causal functionality.

4.3.3 Dwell Time Stimuli Limitations

In the original dwell time slideshow for the *violating expected action* event, the finger waving over the light nearly touched the surface of the light mechanism. I worried that children would interpret this as a non-canonical form of pressing ("swiping") rather than a true violation of expectation (motion activation). Because I sought to investigate the possible influence of expectancy violations on children's exploratory responses—rather than investigating canonical versus non-canonical forms of the same action—I opted to re-record the waving event so that the finger was farther above the surface of the light. Only a handful of participants saw the former version of the violating event during participation (less than 10).

Despite making this change, the fact children in the *violating expected action* group did not show anticipated boundary advantage effects (where dwell times to boundary slides are higher than those to non-boundary slides) indicates that at least a portion of these children failed to pick up on and respond to the underlying structure of the event (for example, by anticipating when the light would turn on and attending carefully to the event-region where it remained on). Rather, the boundary advantage effects indicated that only children who saw the predictable 'press' event were able to systematically deploy their attention to causally informative event regions. This led us to wonder if children who saw the wave registered the causal nature of the action at all (though a little over half of children in this group attempted a wave, indicating that they did). If I were to design a new set of slideshow images for a similar study, I would angle the view differently in order to capture more clearly that the finger passed above the light in the air. I would also use a

contrasting shirt color to allow the action to 'pop out' more, since the slideshow depicted an object with a black base held in front of a person with a black shirt.

Utilizing an object with such a strong visual affordance (and paired consistent action: press) created a likely ceiling effect for pressing as an exploratory play behavior. Likewise, exhibiting the same object with an unexpected action (wave) created a likely floor effect for the waving exploratory play behavior. This was relevant for the proportional play behavior outcome, which was bimodal as a result (i.e., children tended to press to a great extent when they observed the pressing action, and wave to a much lesser extent when they saw the waving action). I attempted to mitigate this bimodality by transforming the play outcome measure when necessary or using analytic approaches that could accommodate the nature of the distribution for this variable (such as the zero-one beta regression). However, I note this as a limitation which could be addressed in future research by including additional visually-interesting object components—such as distractor items—to lessen the likelihood of bimodality in the data.

4.3.4 Computer Mouse Limitations

Using children's mouse clicks as a proxy for their attentional engagement potentially presented a further limitation. Despite using a child friendly mouse that was shaped for a small hand with large buttons (and a bright sticker on top of the left button), some children (especially younger children) had trouble understanding how to make a clicking motion on the mouse with only one finger while leaving the rest of the hand enclosed on top of the mouse. A handful of parents commented that their child had never used a mouse before. This limitation was partially mitigated by the training task, where we guided children in how to click when they were having trouble. However, children's inexperience with using a mouse sometimes impacted their ability to perform clicking—leading to longer or irregular dwell time patterns that may not have been

indicative of their attention, per se. However, the vast majority of children in the sample had no trouble with clicking the mouse. As well, given that dwell time analyses are mainly focused on patterns of dwelling rather than merely between-group mean differences in dwelling, patterns can emerge even when a child clicked slowly.

4.3.5 Operationalization Limitations

Children's accuracy scores on the memory task indicated difficulty in remembering, or possibly even understanding, that they had seen a wave during the dwell time slideshow. In an effort to make the memory questions clear to children, we had printed out still images from the slideshows to ask if children saw a press (while pointing to a still image showing the press action), wave (while pointing to a still image showing the wave action), or both (while pointing back and forth between both still images). However, it is possible that children were unfamiliar with the label 'wave' as applying to this type of action—for example, children may have interpreted the word "wave" to mean the kind of waving one enacts when saying hello or good-bye, rather than the kind of wave exemplified in the present studies' violating expected action slideshow. While a limitation, this issue also points to the strength of using dwell time as an additional, implicit measure of children's engagement with the events depicted in the slideshows.

While collecting data in the museum setting would seem to increase ecological validity relative to lab settings, I note that a one-on-one pedagogical encounter with an adult stranger in a museum may have been quite unique in children's experience. And it is possible that being in an exciting museum environment with many opportunities for discovery may have primed children for exploration. That said, this method of participant recruitment has frequently been used to answer similar research questions (Corriveau et al., 2016).

Children's high level of expectation that the light would be activated by pressing meant that almost all children pressed at least once, and some did nothing other than press the light, even when having witnessed a waving demonstration. This near-ceiling level of pressing may have made it difficult to detect condition differences in exploratory play activities, such as a predicted difference in proportion of time spent with the demonstrated action in the knowledgeable verbal pedagogy and naïve verbal pedagogy conditions. It seems less obvious, but is also plausible, that this ceiling level expectation concerning pressing may have affected dwell-time patterns, in that some children may have had such a strong expectation that pressing would occur that they failed to detect that something other than pressing (i.e., waving) was occurring as the event unfolded during their slide-show viewing in the violated expectation action groups. That is, they may not have noticed the violation of expectation at all, which could have affected the systematicity of dwell-time patterns in the violating expected action groups, as well as having undercut detection of dwell-time differences related to action-violating versus action-consistent expectations. The ceiling effect issue therefore raises the need for confirmation of the findings from this study with other stimuli for which a greater range of variability in expectations and exploratory patterns emerges (as noted in relation to the novel object limitations above).

4.4 Broader Implications

On the whole, Study 1 findings confirmed prior evidence documenting tradeoffs in children's response to varying pedagogical context-setting techniques (e.g., Bonawitz, et al., 2011). In particular, relative to the directive, knowledgeable verbal pedagogy statement ("This is how my toy works"), the inquiry-oriented naive verbal pedagogy statement ("I wonder how this toy works") enhanced children's tendency to actively explore in ways that extended beyond the action demonstration they'd witnessed. Likewise, these results confirmed previous findings regarding the

power of children's *a priori* expectations about an object in shaping the way they actively engage with that object (e.g., Stahl & Feigenson, 2017). Specifically, action demonstrations that violated children's expectations elicited broader exploratory activity relative to demonstrations that merely confirmed what children expected. Yet naïve pedagogy and expectation violations in combination were uniquely powerful in triggering intensified exploratory activity. This novel finding points to knowledge-generating processes such as active, exploratory play arising as joint products of multiple sources of influence on children's processing of unfolding events.

The joint product finding received further emphasis from analyses I undertook investigating links between exploratory play and individual differences in executive skill and temperament. Children with stronger executive skill, extraversion/surgency, and effortful control produced play higher in fidelity to what they'd witnessed, particularly when a demonstration violated their expectations (e.g., waving rather than pressing when waving had been demonstrated). Executive skill also appeared to help children notice causal evidence that their own exploratory efforts gave them the opportunity to witness, such as discovering during play that waving over the light caused the light to turn on, even though pressing was the action they'd seen demonstrated. These findings underscore further that multiple sources of influence jointly shape children's knowledge-updating processes. They also provide an altogether new window on the complex interplay of many factors as children actively work – or more accurately, actively play – to make sense of the causal structure of things.

I included the dwell-time paradigm in the dissertation (Study 2) to supplement exploratory play as a source of evidence about children's knowledge-updating processes. Specifically, dwell time patterns offered evidence concerning children's active information-sampling as they experienced the unfolding action demonstrations that supplied their first available evidence

regarding the light's causal profile. First, analyses revealed a classic dwell-time signature of causal understanding in children's attentional patterns (Kosie & Baldwin, 2019, 2021); that is, children displayed enhanced dwelling to boundary regions of the consistent expected action (pressing) demonstration (regions where one action transitioned to the next and resulted in causal effects) relative to non-boundary regions. In contrast, however, a boundary advantage was not evident in children's dwelling patterns as they viewed the waving demonstration that violated their expectations. On the one hand, this boundary-advantage-differential based on action expectations indicates that children were responding to differences within the consistent and violating events. The analyses made clear that this differential was not due merely to the fact that there were small surface motion differences between the consistent (pressing) and violating (waving) actions, given that these analyses controlled for surface motion differences via the pixel change measure. The boundary advantage differential to violating versus consistent events thus seems likely to reflect differences in children's conceptual processing. On the conceptual level, children may have tended to detect the violation of their expectations while at the same time not quite achieving a causal analysis of the waving action being demonstrated, leaving them unable to upregulate attentional sampling at boundary regions they couldn't identify. I explored this speculative analysis of the dwell-time findings further by isolating dwell-time patterns based on whether children subsequently displayed greater exploration– waving and performing unique actions– during the play period, versus those who did not (in relation to executive function skills). My rationale was that enhanced exploratory play served as evidence that children had achieved a causal analysis of the waving action. These analyses revealed evidence that both boundary advantage and overall dwell times were affiliated with these exploratory behaviors via EF. Although intriguing, these

findings should be treated with particular caution given the highly exploratory nature of the analyses and the fact that sample sizes in these comparison were underpowered.

4.5 Future Directions

In Study 2, my primary focus was understanding the relationship between dwell time and the primary independent variables (verbal pedagogy and expected action groups). Therefore, I undertook an analysis on the direct relationship between dwell times, play behavior, and EF only in an exploratory sense. I found that exploratory behavior for lower EF children was associated with a boundary advantage, and for higher EF children, performing the action demonstrated was associated with overall shorter dwell times.

In the analyses just reported, children's play behavior was treated as the anchor; the question of interest was which dwell-time patterns were associated with these anchoring play patterns. One area of open interest concerns the opposite directionality: which dwell time patterns during slide-show viewing predict children's subsequent play patterns? There is the potential that a dwell-time "signature" of causal processing might bear systematic relations to children's subsequent exploratory behavior. This could allow us to tailor pedagogical strategies to children's individual learning needs based upon what their attentional signatures tell us about their level of causal understanding. Therefore, creating a corpus of children's dwell times related to causal learning events is a larger research goal that extends beyond this project, with the potential for allowing us to understand more about the relationship between dwell time and exploratory behaviors.

Combining measurement of dwell-time patterns with exploratory play offers myriad possibilities for investigation of knowledge-updating processes. To play out just one example, it should be possible to use this combination of measures to gain evidence of the extent to which exploratory play reflects children's hypothesis-driven interventions on the world, by which they

generate new evidence to inform and update their beliefs. One could measure dwell-time patterns to an event such as the expectation-violating wave action on the light, then offer children the light to play with and measure exploratory play activities, while following the play period up by giving children another opportunity to dwell through the waving action. Of interest would be the extent to which particular reorganization in dwell-time patterns on second versus first viewing depend on the nature of the exploratory activities that children engaged in after first viewing. Some interesting control comparison groups might be a) when, instead of playing with the light themselves, children simply watch someone else play with the light, or b) children are given an entirely different (causally unrelated) object to explore. Findings from such research would help to illuminate the extent to which children's active, exploratory play reorganizes their knowledge-related processing of events they encounter, and I hope to one day undertake research of this kind.

4.6 Conclusion

In the dissertation, children who heard naïve verbal pedagogy were more exploratory and showed greater differences in their attention to event demonstrations upon the second viewing of an event compared to those who heard knowledgeable verbal pedagogy. Further, when children both heard naïve verbal pedagogy and saw an event demonstration that violated their expectations, they were especially exploratory compared to other groups. Children's executive function (EF) skills also related to their play behaviors, with those higher in EF especially likely to a) perform the violating action and b) perform a greater number of exploratory actions than those lower in EF. Put another way, EF played a moderating role, with children who had stronger EF skills showing greater sensitivity to pedagogical framing in contexts of expectation violation, whether expressed as intensified fidelity under knowledgeable framing or expanded breadth under naïve framing. Combining the dwell-time paradigm with observation of children's exploratory play activities provided new insight. Together, these looking time and exploratory play techniques shed light on

how visual sampling of events is shaped by expectations, changes as action evidence emerges, and ultimately alters the way that children engage with the world during play.

These studies illuminate how children respond in learning situations when evidence varies across multiple domains—socio-linguistic and direct causal action— much as they confront such complexity in their everyday lives. Children take in complex, multilayered streams of evidence and bring them jointly to bear as they update their inferences. These findings illustrate in a new way that children, to varying degrees in relation to their regulatory skills, tend to be active, strategic learners, flexibly applying different forms of pedagogy to their learning while remaining sensitive to the details of the event streams unfolding across time as they observe new evidence. Like anglers, learners take in the stream of unfolding evidence and design their casting to optimize the chances of snagging the most informative insights.

APPENDIX A

Parent-report Demographic Questionnaire

Q1 What is the participating child's gender?

- ○ Male
- ○ Female
- ○ Non-binary/third gender
- ○ Other (please specify)_____
- ○ Prefer not to say

Q2 What is the participating child's age?

- ○ 2 years (24-35 months)
- ○ 3 years (36-47 months)
- ○ 4 years (48-59 months)
- ○ 5 years (60-71 months)
- ○ 6 years (72-83 months)
- ○ 7 years (84-95 months)
- ○ 8 years (96-107 months)
- ○ Other (please specify)_____

Q3 What is the participating child's ethnicity?

- ○ Hispanic or Latino
- ○ Not Hispanic or Latino

Q4 What is the participating child's race?

- ○ Alaska Native or Indigenous American
- ○ Asian
- ○ Biracial (please specify)
- ○ Black of African American
- ○ Native Hawaiian or Pacific Islander
- ○ White
- ○ Prefer not to say
- ○ Other (please specify)_____

Q5 Think of this ladder as showing where people stand in their communities. People define community in different ways. Please define it in whatever way is most meaningful to you. At the top of the ladder are the people who have the most standing in their community.



Where would you put yourself on this ladder (select one)?

- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5
- ○ 6
- ○ 7

- ○ 8
- ○ 9
- ○ 10

Q6 Think of this ladder as showing where people stand in the United States. At the top of the ladder are the people who are best off—those who have the most money, the most education, and the most respected jobs. At the bottom are the people who are the worst off—who have the least money, least education, and the least respected job or no job. The higher up you are on this ladder, the closer you are to the people at the very top; the lower you are, the closer you are to the people at the very bottom.



Where would you place yourself on this ladder (select one)?

- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5
- ○ 6
- ○ 7
- ○ 8
- ○ 9
- ○ 10

Q7 What is the highest level of education that you have completed as of today?

- ○ 7th Grade or earlier
- ○ Junior High (9th grade)
- ○ Some High School
- ○ High School Graduate or GED
- ○ Some college credit but less than 1 year
- ○ One or more years of college at a 2-year program, no degree
- ○ One year of college at a 4-year program, no degree
- ○ Two years of college at a 4-year program, no degree
- ○ Three years or more of college at a 4-year program, no degree
- ○ Associate's degree (e.g. AA, AS)
- ○ Bachelor's degree (e.g. BA, BS)
- ○ Master's degree (e.g. MA, MS)
- ○ Professional degree (e.g. MD, DDS, JD)
- ○ Doctorate (e.g. PhD EdD)
- ○ Other (please specify)_____

APPENDIX B

Children's Behavior Questionnaire Short Form (items included for: *Approach, Activity Level, Attentional Focusing, Impulsivity, Inhibitory Control*, and *Shyness*

1	2	3	4	5	6	7	NA
extremely untrue	quite untrue	slightly untrue	neither true nor	slightly true	quite true	extremely true	not applicable

In the past 6 months, my child:

1. Seems always in a big hurry to get from one place to another.

   1 2 3 4 5 6 7 NA

2. Gets so worked up before an exciting event that s/he has trouble sitting still.

   1 2 3 4 5 6 7 NA

3. Usually rushes into an activity without thinking about it.

   1 2 3 4 5 6 7 NA

4. Seems to be at ease with almost any person.

   1 2 3 4 5 6 7 NA

5. Tends to run rather than walk from room to room.

   1 2 3 4 5 6 7 NA

6. Gets very enthusiastic about the things they do.

   1 2 3 4 5 6 7 NA

7. When practicing an activity, has a hard time keeping their mind on it.

   1 2 3 4 5 6 7 NA

8. When outside, often sits quietly.

   1 2 3 4 5 6 7 NA

9. Will move from one task to another without completing any of them.

   1 2 3 4 5 6 7 NA

10. Moves about actively (runs, climbs, jumps) when playing in the house.

    1 2 3 4 5 6 7 NA

11. Often rushes into new situations.

    1 2 3 4 5 6 7 NA

12. Takes a long time in approaching new situations.

    1 2 3 4 5 6 7 NA

13. Is sometimes shy even around people they have known a long time.

1 2 3 4 5 6 7 NA

14. Can wait before entering into new activities if they are asked to.

1 2 3 4 5 6 7 NA

15. Sometimes seems nervous when talking to adults they have just met.

1 2 3 4 5 6 7 NA

16. Is slow and unhurried in deciding what to do next.

1 2 3 4 5 6 7 NA

17. Prepares for trips and outings by planning things they will need.

1 2 3 4 5 6 7 NA

18. Becomes very excited while planning for trips.

1 2 3 4 5 6 7 NA

19. Prefers quiet activities to active games.

1 2 3 4 5 6 7 NA

20. Tends to say the first thing that comes to mind, without stopping to think about it.

1 2 3 4 5 6 7 NA

21. Acts shy around new people.

1 2 3 4 5 6 7 NA

22. Has trouble sitting still when s/he is told to (at movies, church, etc.).

1 2 3 4 5 6 7 NA

23. Becomes very excited before an outing (e.g., picnic, party).

1 2 3 4 5 6 7 NA

24. Is comfortable asking other children to play.

1 2 3 4 5 6 7 NA

25. When drawing or coloring in a book, shows strong concentration.

1 2 3 4 5 6 7 NA

26. Is good at following instructions.

1 2 3 4 5 6 7 NA

27. Sometimes turns away shyly from new acquaintances.

1 2 3 4 5 6 7 NA

28. When building or putting something together, becomes very involved in what they are doing, and works for long periods.

1 2 3 4 5 6 7 NA

29. Approaches places they have been told are dangerous slowly and cautiously.

1 2 3 4 5 6 7 NA

30. Can easily stop an activity when they are told "no."

1 2 3 4 5 6 7 NA

31. Is among the last children to try out a new activity.

1 2 3 4 5 6 7 NA

32. Is easily distracted when listening to a story.

1 2 3 4 5 6 7 NA

33. Is full of energy, even in the evening.

1 2 3 4 5 6 7 NA

34. Sometimes becomes absorbed in a picture book and looks at it for a long time.

1 2 3 4 5 6 7 NA

35. Remains pretty calm about upcoming desserts like ice cream.

1 2 3 4 5 6 7 NA

36. Looks forward to family outings, but does not get too excited about them.

1 2 3 4 5 6 7 NA

37. Likes to sit quietly and watch people do things.

1 2 3 4 5 6 7 NA

APPENDIX C

Head-Toes-Knees-Shoulders Task Researcher script in italics and Scoring Sheet

Now we're going to play a game. The game has two parts. First, I want you to copy what I do. Touch your head. Wait for the child to put BOTH his/her hands on head.

Good! Now touch your toes. Wait for the child to put his/her hands on toes.

Good! Repeat the two commands with motions again, or until the child imitates you correctly: *Touch your head. Touch your toes.*

Now we're going to be a little silly and do the opposite of what I say. When I say to touch your head, instead of touching your head, you touch your toes. When I say to touch your toes, you touch your head. So, you're doing something different from what I say.

Now we're going to play a game. The game has two parts. First, I want you to copy what I do. Touch your head. Wait for the child to put BOTH his/her hands on head.

Good! Now touch your toes. Wait for the child to put his/her hands on toes.

Good! Repeat the two commands with motions again, or until the child imitates you correctly.*) Touch your head. Touch your toes.*

Now we're going to be a little silly and do the opposite of what I say. When I say to touch your head, instead of touching your head, you touch your toes. When I say to touch your toes, you touch your head. So, you're doing something different from what I say.

PART I TRAINING: (circle child's response)

If retraining is needed, repeat A1 (record both scores). If no retraining is needed, only record initial score (top)

A1. *What do you do if I say "touch your head"?*		
0 (head)	1	2 (toes)
0 (head)	1	2 (toes)

If retraining is needed, repeat A2 (record both scores). If no retraining is needed, only record initial score (top)

A2. *What do you do if I say "touch your toes"?*		
0 (toes)	1	2 (head)
0 (toes)	1	2 (head)

Retraining statement: *Remember, when I say to touch your toes, you touch your head, so you are doing something different from what I say. Let's try again.*

PART I PRACTICE: (circle child's response)

	Incorrect	Self-Correct	Correct
B1. *Touch your head*	0 (head)	1	2 (toes)
B2. *Touch your toes*	0 (toes)	1	2 (head)
B3. *Touch your head*	0 (head)	1	2 (toes)
B4. *Touch your toes*	0 (toes)	1	2 (head)
If retraining is needed, repeat 4 Practice items (record all 8 scores). If no retraining is needed, only record initial scores (top)			
B1. *Touch your head*	0 (head)	1	2 (toes)
B2. *Touch your toes*	0 (toes)	1	2 (head)
B3. *Touch your head*	0 (head)	1	2 (toes)
B4. *Touch your toes*	0 (toes)	1	2 (head)

1. *Touch your head*	0 (head)	1	2 (toes)
2. *Touch your toes*	0 (toes)	1	2 (head)
3. *Touch your toes*	0 (toes)	1	2 (head)
4. *Touch your head*	0 (head)	1	2 (toes)
5. *Touch your toes*	0 (toes)	1	2 (head)
6. *Touch your head*	0 (head)	1	2 (toes)
7. *Touch your head*	0 (head)	1	2 (toes)
8. *Touch your toes*	0 (toes)	1	2 (head)
9. *Touch your head*	0 (head)	1	2 (toes)
10. *Touch your toes*	0 (toes)	1	2 (head)

Administer Part II if child responds correctly to 5 or more items on Part I of the task, or if the child is in kindergarten or beyond.

Ok, now that you've got that part, we're going to add a part. Now, you're going to touch your shoulders and your knees. First, touch your shoulders. Touch your shoulders; wait for the child to touch his/her shoulders with both hands.

Now, touch your knees. Touch your knees; wait for the child to touch his/her knees with both hands. Repeat with four alternating commands (no demo) until the child has imitated you correctly or it is clear the child does not comprehend the task. *Touch your shoulders. Touch your knees. Touch your shoulders. Touch your knees.*

Ok, now we're going to be silly again. You're going to keep doing the opposite of what I say like before. But this time, you're going to touch your knees and shoulders. When I say to touch your knees, you touch your shoulders, and when I say to touch your shoulders, you touch your knees.

PART II TRAINING: (circle child's response):

If retraining is needed, repeat A2 (record both scores). If no retraining is needed, only record initial score (top)

C1. *What do you do if I say "touch your knees"?*		
0 (knees)	1	2 (shoulders)
0 (knees)	1	2 (shoulders)

PART II PRACTICE (circle child's response):

	Incorrect	Self-Correct	Correct
D1. *Touch your knees*	0 (knees)	1	2 (shoulders)
D2**.** *Touch your shoulders*	0 (shoulders)	1	2 (knees)
D3. *Touch your knees*	0 (knees)	1	2 (shoulders)
D4**.** *Touch your shoulders*	0 (shoulders)	1	2 (knees)
If retraining is needed, repeat 4 Practice items (record all 8 scores). If no retraining is needed, only record initial scores (top)			
D1. *Touch your knees*	0 (knees)	1	2 (shoulders)
D2**.** *Touch your shoulders*	0 (shoulders)	1	2 (knees)
D3. *Touch your knees*	0 (knees)	1	2 (shoulders)
D4**.** *Touch your shoulders*	0 (shoulders)	1	2 (knees)

PART II TEST (circle child's response)

11. *Touch your head*	0 (head)	1	2 (toes)
12. *Touch your toes*	0 (toes)	1	2 (head)
13. *Touch your knees*	0 (knees)	1	2 (shoulders)
14. *Touch your toes*	0 (toes)	1	2 (head)
15. *Touch your shoulders*	0 (shoulders)	1	2 (knees)
16. *Touch your head*	0 (head)	1	2 (toes)
17. *Touch your knees*	0 (knees)	1	2 (shoulders)
18. *Touch your knees*	0 (knees)	1	2 (shoulder)
19. *Touch your shoulders*	0 (shoulders)	1	2 (knees)
20. *Touch your toes*	0 (toes)	1	2 (head)

APPENDIX D

Preregistration Deviation Table (strategy based on Willroth & Atherton, 2024)

Deviation Type	Description	Impact of Deviation in Dissertation
Study Design	The "Expected Action" variable described throughout the dissertation was named "Visual Affordance" in the pre-registered plan. The pre-registration label "Low Visual Affordance" was instead termed "Violating Expected Action" in the dissertation, and the pre-registered variable "High Visual Affordance was instead termed "Consistent Expected Action" in the dissertation.	Rather than describing the action manipulation based on what the stimulus object visually afforded, the manipulation in the dissertation was instead framed around children's expectations for action on the stimulus object.
Exclusion criteria	No details were included in the pre-registration about age-related exclusion; however, in the dissertation, 3 participants were excluded for being under 3 years or over 7 years old. Additionally, no details were included in the pre-registration about excluding children on the basis of no parental consent for video recording; however, 3 participants were excluded from play analyses in Study 1 for this reason.	Six participants were excluded beyond the criteria listed in the pre-registration on the basis of constricting the number of discrete year age groups included, and on the basis of having no video data necessary for coding children's play behaviors (due to lack of parental consent).
Research questions	Hypotheses, rather than research questions, were pre-registered in accordance with the OSF template at the time of access. Research questions related to those hypotheses are listed in the introductions and results sections for Studies 1 and 2.	While the research questions listed in Studies 1 and 2 in the dissertation reflect the pre-registered hypotheses, they are not identically matched to those hypotheses. For example, some research questions are phrased in open-ended ways that don't map perfectly to the way the hypotheses in the pre-registration were stated. This does not change the overall interpretation or reporting of the results but makes the pre-registered hypotheses less relevant to the study outcomes.

Hypotheses	A number of pre-registered exploratory hypotheses were not addressed in the dissertation.	The exploratory hypotheses listed in the pre-registration focused on specific predictions related to particular dwell time patterns (e.g., proactive and reactive dwell time patterns) which were not discussed in the dissertation. However, the exploratory reporting in the dissertation aligns, in a general sense, with the pre-registered exploratory analysis plan.
Covariates	In the pre-registration, examining relations between main outcomes and executive function was listed in the exploratory analysis plan. However, no details were included in the pre-registration about controlling for age while examining links between executive function and the main outcomes of interest. Additionally, the pre-registration included a plan to create an SES index based upon parent education and parent Subjective Social Status scores. However, while this index was created, it was never reported nor included in analyses for the dissertation.	After submitting the pre-registration, we realized that controlling for age while examining links between executive function and other variables was necessary. Additionally, we opted not to include an SES measure in dissertation results given that examining links between SES and other variables was not a focus of Study 1 or Study 2.
Inferential criteria	The pre-registration stated the planned use of a Holm correction for p-values, but a Bonferroni correction was used in the dissertation.	We opted to use a Bonferroni correction rather than Holm for consistency across reported results (i.e., the same cutoff is applied throughout rather than the stepwise comparisons used for Holm), along with the desire to be conservative in estimates, especially for 3-way interactions where the full sample is under 200 participants.

REFERENCES

Adler, N. E., Epel, E. S., Castellazzo, G., & Ickovics, J. R. (2000). *Subjective SES Scale* [Database record]. APA PsycTests. https://doi.org/10.1037/t20985-000

Anderson, M. M., Kiverstein, J., Miller, M. & Roepstoff, A. (2023). Play in predictive minds: A cognitive theory of play. Psychological Review, 130(2), 462-479. https://doi.org/10.1037/rev0000369

Bach, J. & Pauen, S. (2025). Are children more likely to copy parents at home or experimenters in the lab? Developmental changes in overimitation between 4 and 7 years of age. *Developmental Psychology,* Advance online publication. https://doi.org/10.1037/dev0002007

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

Baldwin D. A., Baird J. A. (2001). Discerning intentions in dynamic human action. *Trends in Cognitive Sciences*, 5, 171–178. https://doi.org/10.1016/S1364-6613(00)01615-6

Baldwin, D. A., & Kosie, J. E. (2021). How does the mind render streaming experiences as events? *Topics in Cognitive Science, 13*(1), 79-105. https://doi.org/10.1111/tops.12502

Baldwin, D., Loucks, J. & Sabbagh, M., (2008). Pragmatics of human action. T. F. Shipley & J. Zacks (Eds.), *Understanding events: From perception to action*. Oxford University Press.

Bass, I. & Bonawitz, E. (2024). Early environments and exploration in the preschool years. *PLoS ONE, 19*(6): e e0305353. https://doi.org/ 10.1371/journal.pone.0305353

Bauer, J. & Booth, A. E. (2019). Exploring potential cognitive foundations of scientific literacy in preschoolers: Causal reasoning and executive function. Early *Childhood Research Quarterly, 46,* 275-284. https://doi.org/10.1016/j.ecresq.2018.09.007

Blair, C. (2018). Executive function and early childhood education. *Current Opinions in Behavioral Science, 10,* 102-107. 10.1016/j.cobeha.2016.05.009

Bonawitz, E., Shafto, P., Gweon, H., Goodman, N., Spelke, E., & Schulz, L. (2011) The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition, 120*(3): 322-30. https://doi.org/10.1016/j.cognition.2010.10.001

Bonawitz, E. & Ullman, T. (2024). Bayesian models of cognitive development. In T. L. Griffiths, N. Chater & J. B. Tenenbaum (Eds.). *Bayesian Models of Cognition* (pp. 499-517). The MIT Press. https://mitpress.ublish.com/ebook/bayesian-models-of-cognition-reverse-engineering-the-mind-preview/12799/499

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

Brezack, N., Radovanovic, M. & Woodward, A. L. (2021). Everyday interactions support toddlers' learning of conventional actions on artifacts. *Journal of Experimental Child Psychology, 210,* e105201. https://doi.org/10.1016/j.jecp.2021.105201

Brod, G., Breitwieser, J., Hasselhorn, M. & Bunge, S. A. (2019). Being proven wrong elicits learning in children – but only in those with higher executive function skills. *Developmental Science, 23*, e12916. https://doi.org/10.1111/desc.12916

Brooks ME, Kristensen K, van Benthem KJ, Magnusson A, Berg CW, Nielsen A, Skaug HJ, Maechler M, Bolker BM (2017). "glmmTMB Balances Speed and Flexibility Among Packages for Zero-Inflated Generalized Linear Mixed Modeling." *The R Journal*, 9(2), 378–400. doi:10.32614/RJ-2017-066.

Brosseau-Liard, P., Cassels, T., & Birch, S. (2014). You seem certain but you were wrong before: Developmental change in preschoolers' relative trust in Accurate versus confident speakers. *PLoS ONE, 9*(9). https://doi.org/10.1371/journal.pone.0108308

Buchsbaum, D., Griffiths, T. L., Gopnik, A., & Baldwin, D. (2009). Learning from actions and their consequences: Inferring causal variables from continuous sequences of human action. *Proceedings of the 31st annual conference of the Cognitive Science Society* (Vol. 103, pp. 134). Cognitive Science Society Austin, TX.

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6), 679-698. https://doi.org/10.1109/TPAMI.1986.4767851

Carey, S. Zaitchik, D. & Bascandziev, I. (2015). Theories of development: In dialog with Jean Piaget. *Developmental Review, 38*, 36-54. https://doi.org/10.1016/j.dr.2015.07.003

Carlson, S. M., Davis, A. C. & Leach, J. G. (2005). Less is more: executive function and symbolic representation in preschool children. *Psychological Science, 16*(8). 10.1111/j.1467-9280.2005.01583.x

Champely S (2023). *pwr: Basic Functions for Power Analysis*. R package version 1.3-0, https://github.com/heliosdrm/pwr.

Chisholm, R. M. (1957). *Perceiving: A philosophical study.* Cornell University Press. https://philpapers.org/rec/CHIPAP

Chong, S. Y., Chittleborough, C. R., Gregory, T., Lynch, J., Mittinty, M. & Smithers, L. G. (2019). The controlled direct effect of temperament at 2-3 years on cognitive and academic outcomes at 6-7 years. *PLoS ONE 14*(6): e0204189. https://doi.org/10.1371/journal.pone.0204189

Colantonio, J., Bascandziev, I., Theobald, M., Brod, G. & Bonawitz, E. (2023). Seeing the error in my *"Bayes"*: A quantified degree of belief change correlates with children's pupillary surprise responses following explicit predictions. *Entroy, 25,* 211. https://doi.org/10.3390/e25020211

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

Corriveau, K. H., Kipling, R., Ronfard, S., Biarnes, M. C., Jeye, B. M. & Harris, P. L. (2016). The Living Laboratory® Model: A mutual professional development model for museum-based research partnerships. In D. M. Sobel & J. L. Jipson (2016). *Cognitive development in museum settings: related research and practice* (pp. 65-83). Psychology Press.

Corriveau, K. H., & Kurkul, K. E. (2014). "Why does rain fall?": Children prefer to learn from an informant who uses noncircular explanations. *Child Development*, *85*(5), 1827–1835. https://doi.org/10.1111/cdev.12240

Csibra, G. & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Science, 13*(4), 148-53. 10.1016/j.tics.2009.01.005

Dekker, M. C., Ziermans, T. B., Spruijt, A. M., & Swaab, H. (2017). Cognitive, parent and teacher rating measures of executive functioning: Shared and unique influences on school achievement. *Frontiers in Psychology*, *8*, 48. https://doi.org/10.3389/fpsyg.2017.00048

Doan, T., Castro, A., Bonawitz, E. & Denison, S. (2020). "Wow, I did it!" Unexpected success increases preschoolers' exploratory play on a later task. *Cognitive Development, 55*, 100925.https://doi.org/10.1016/j.cogdev.2020.100925

Fedra, E. & Schmidt M. F. H. (2019). Older (but not younger) preschoolers reject incorrect knowledge claims. *British Journal of Developmental Psychology, 37,* 130-145. https://doi.org/0.1111/bjdp.12264

Fei, X. (2019). Towards a rational constructivist theory of cognitive development. *Psychological Review, 126*(6), 841-864. https://doi.org/10.1037/rev0000153

Frazier, B. N., Gelman, S. A. & Wellman, H. M. (2010). Preschoolers' search for explanatory information within adult-child conversation. *Child Development, 80*(6), 1592-1611. 10.1111/j.1467-8624.2009.01356.x

Gaspelin, N. & Luck, S. J. (2019). The role of inhibition in avoiding distraction by salient stimuli. *Trends in Cognitive Science, 22*(1). 10.1016/j.tics.2017.11.001

Gergely, G., Bekkering, H. & Király, I. (2002). Rational imitation in preverbal infants. *Nature, 415*, 755-6. https://doi.org/10.1038/415755a

Gettier, E. L. (1963). Is justified true belief knowledge? *Analysis, 23*(6), 121-123.

Goldberg, S. (2008). Testimonial knowledge in early childhood revisited. *Philosophy and Phenomenological Research, 76*(1), 1–36. https://www.jstor.org/stable/40041150

Gopnik, A. (2020). Childhood as a solution to explore-exploit tensions. *Philosophical Transactions R. Soc. B,* 375. http://dx.doi.org/10.1098/rstb.2019.0502

Gopnik, A. & Meltzoff, A. N. (1997). The scientist as child. In A. Gopnik & A. N. Meltzoff (Eds.). *Words, thoughts, and theories* (pp. 13-47). The MIT Press. https://doi.org/10.7551/mitpress/7289.003.0005

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. https://www.jstor.org/stable/1131970

Gopnik, A. & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms and the theory theory. *Psychological Bulletin, 138*(6), 1085-1108. https://doi.org/10.1037/a0028044

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

Grice, H. P. (1975). Speech acts. In *Syntax and semantics.* In P. Cole & J. L. Morgan (pp. 41-58). Academic Press, Inc. https://philpapers.org/rec/GRIISA

Gweon, H. Pelton, H., Konopka, J. A. & Schulz, L. E. (2014). Signs of omission: Children selectively explore when teachers are under-informative. *Cognition, 132*, 335-341. http://dx.doi.org/10.1016/j.cognition.2014.04.013

Hard, B. M., Meyer, M., & Baldwin, D. (2019). Attention reorganizes as structure is detected in dynamic action. *Memory & Cognition, 47*(1), 17–32. 10.3758/s13421-018-0847-z

Hard, B. M., Recchia, G., & Tversky, B. (2011). The shape of action. *Journal of Experimental Psychology: General, 140*, 586– 604. https://doi.org/10.1037/a0024310

Harris, P. L., Koenig, M. A., Corriveau, K. H., & Jaswal, V. K. (2018). Cognitive foundations of learning from testimony. *Annual Review of Psychology, 69*, 251-273. https://doi.org/10.1146/annurev-psych-122216-011710

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences, 7*(11), 498-504. https://doi.org/10.1016/j.tics.2003.09.006

Horowitz, A. C. & Frank, M. C. (2016). Children's pragmatic inferences as a route for learning about the world. *Child Development, 87*(3), 807-819. https://doi.org/10.1111/cdev.12527

Eycke, K. D. T. & Dewey, D. (2016). Parent-report and performance-based measures of executive function assess different constructs. *Child Neuropsychology*, *22*(8), 889–906. https://doi.org/10.1080/09297049.2015.1065961

Jaswal, V. K., Croft, A. C., Setia, A. R., & Cole, C. A. (2010). Young children have a specific, highly robust bias to trust testimony. *Psychological Science, 21*(10), 1541–1547. https://doi.org/10.1177/0956797610383438

Jaswal, V. K., Pérez-Edgar, K., Kondrad, R. L., Palmquist, C. M., Cole, C. A., & Cole, C. E. (2014). Can't stop believing: Inhibitory control and resistance to misleading testimony. *Developmental Science, 17*(6), 965– 976. https://doi.org/10.1111/desc.12187

Keupp, S., Behne, T. & Rakoczy, H. (2018). The rationality of (over)imitation. *Perspectives on Psychological Science, 13*(6), 678-687. https://doi.org/10.1177/1745691618794921

Kooiker, M. J. G., van der Steen, J. & Pel, J. J. M. (2016) Development of salience-driven and visually-guided eye movement responses. *Journal of Vision, 16*(5): 18, 1-11. https://doi.org/10.1167/16.5.18

Kosie, J. E., & Baldwin, D. (2019). Attention rapidly reorganizes to naturally occurring structure in a novel activity sequence. *Cognition, 182*, 31– 44. https://doi.org/10.1016/j.cognition.2018.09.004

Kosie, J. E., & Baldwin, D. A. (2021). Dwell times showcase how goal structure informs preschoolers' analysis of unfolding motion patterns. *Child Development, 92*(6). https://doi.org/10.1111/cdev.13661

Kuznetsova A, Brockhoff PB, Christensen RHB (2017). "lmerTest Package: Tests in Linear Mixed Effects Models." *Journal of Statistical Software*, *82*(13), 1–26. doi:10.18637/jss.v082.i13.

Lapidow, E. & Walker, C. M. (2020). Informative experimentation in intuitive science: Children select and learn from their own causal interventions. *Cognition, 201*, 104315. https://doi.org/10.1016/j.cognition.2020.104315

Legare, C. H., Gelman, S. A. & Wellman, H. M. (2010). Inconsistency with prior knowledge triggers children's causal explanatory reasoning. *Child Development, 81*(3), 929-944. 10.1111/j.1467-8624.2010.01443.x

Lenth, R. V. (2024). Emmeans: Estimated marginal means, aka least-squares means (R package version 4.5.1). https://CRAN.R-project.org/package=emmeans

Liquin, E. G. & Gopnik, A. (2022). Children are more exploratory and learn more than adults in an approach-avoid task. Cognition, *218*, 104940. https://doi.org/10.1016/j.cognition.2021.104940

Loucks, J., & Baldwin, D. (2009). Sources of information for discriminating dynamic human actions. *Cognition, 111*, 84– 97. https://doi.org/10.1016/j.cogni tion.2008.12.010

Margoni, F., Surian, L., & Baillargeon, R. (2024). The violation-of-expectation paradigm: A conceptual overview. *Psychological Review, 131*(3), 716-748.

Mazzarella, D. & Pouscoulous, N. (2021). Pragmatics and epistemic vigilance: A developmental perspective, *Mind & Language, 36,* 355-376. https://doi.org/10.1111/mila.12287

Meyer, M., Baldwin, D. A., & Sage, K. D. (2011). Assessing young children's hierarchical action segmentation. In Proceedings of the 33rd annual conference of the Cognitive Science Society. https://escholarship.org/uc/item/2vp3c8d1

Meltzoff, A. N. (1988). Infant imitation and memory: Nine-month-olds in immediate and deferred tests. *Child Development, 59*(1), 217-225. https://doi.org/10.1111/j.1467-8624.1988.tb03210.x

Miller, P. H. (2010). Piaget's theory: Past, present, and future. In Goswami, U. (Ed.). *The Wiley-Blackwell handbook of childhood cognitive development* (pp. 649-672). John Wiley & Sons, Inc.

Moses, L.J., Carlson, S.M., & Sabbagh, M.A. (2005). On the specificity of the relation between executive function and children's theories of mind. In W. Schneider, R. Schuhmann-Hengsteler & B. Sodian (Eds.), *Young children's cognitive development: Interrelationships among executive functioning, working memory, verbal ability, and theory of mind* (pp. 131–145). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Nagar, I., & Gupta, A. (2017). Examining discrepancy between performance on traditional measures and parent report measure of executive functioning in autism spectrum disorder. *Indian Journal of Psychological Medicine*, *39*(6), 813-816. https://doi.org/10.4103/IJPSYM.IJPSYM_24_17

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology, 28*(1), 28–38. https://doi.org/10.1037/h0035584

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J. (2019). PsychoPy2: experiments in behavior made easy. *Behavior Research Methods.* https://.org/10.3758/s13428-018-01193-y

Perfors, A. & Navarro, D. (2012). What Bayesian modelling can tell us about statistical learning: What it requires and why it works. In P. Rebuschat & J. N. Williams (Eds.). *Statistical learning and language acquisition* (pp. 383-408). De Gruyter, Inc. https://doi.org/10.1016/j.cognition.2010.11.015

Perner, J. (1993). *Understanding the Representational Mind*. The MIT Press. https://doi.org/10.7551/mitpress/6988.001.0001

Piaget, J. (1954) *The Construction of Reality in the Child.* New York: Basic. http://dx.doi.org/10.1037/11168-000

Ponitz, C. C., McClelland, M. M., Mathews, J. S. & Morrison, F. J. (2009). A structured observation of behavioral self-regulation and its contribution to kindergarten outcomes. *Developmental Psychology, 45*(3), 603-619. https://doi.org/10.1037/a0015365

Putnam, S. P., & Rothbart, M. K. (2006). Development of short and very short forms of the Children's Behavior Questionnaire. *Journal of Personality Assessment, 87*(1), 102-12. https://doi.org/10.1207/s15327752jpa8701_09

R Core Team (2024). R: a language environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/

Ross, R. A., & Baldwin, D. A. (2015). Event processing as an executive enterprise. In R. A. Scott, & S. M. Kosslyn (Eds.), Emerging trends in the social and behavioral sciences: An interdisciplinary, searchable, and linkable resource (pp. 1– 13). https://doiorg/10.1002/9781118900772

Rothbart, M. K. (1996). *Children's Behavior Questionnaire (CBQ).* [Database record]. APA PsycTests. https://doi.org/10.1037/t05150-000

Ruggeri, A., Stanciu, O., Pelz, M., Gopnik, A. & Schulz, E. (2023). Preschoolers search longer when there is more information to be gained. *Developmental Science, 27*(1), e13411. https://doi.org/10.1111/desc.13411

Sage, K. D. & Baldwin, D. (2014). Looking to the hands: Where we dwell in complex manual sequences. *Visual Cognition, 22*(8), 1092-1104. http://dx.doi.org/10.1080/13506285.2014.962123

Schneidman, L., Gweon, H., Schulz, L., & Woodward, A. (2016). Learning from others and spontaneous exploration: A cross-cultural investigation. *Child Development, 87*(3): 723-735. https://doi.org/10.1111/cdev.12502

Schulz, L. E., & Bonawitz, E. B. (2007). Serious fun: Preschoolers engage in more exploratory play when evidence is confounded. *Developmental Psychology, 43*(4) https://doi.org/10.1037/0012-1649.43.4.1045

Schulz, L. E., Standing, H. R., & Bonawitz, E. B. (2008). Word, thought, and deed: The role of object categories in children's inductive inferences and exploratory play. *Developmental Psychology, 44*(5), 1266–1276. https://doi-org.uoregon.idm.oclc.org/10.1037/0012-1649.44.5.1266

Şen, H. H., Kiefer, S. L., Aksu, E. & Lucca, K. (2024). Developmental differences in children and adults' enforcement of explore versus exploit search strategies in the United States and Turkey. *Developmental Science, 27*, e13520. https://doi.org/10.1111/desc.13520

Silver, C. H. (2014). Sources of data about children's executive functioning: Review and commentary. *Child Neuropsychology, 20*(1), 1-13. https://doi.org/10.1080/09297049.2012.727793

Sim, Z.L. and Xu, F. (2017), Infants preferentially approach and explore the unexpected. *Br J Dev Psychol, 35*: 596-608. https://doi-org.uoregon.idm.oclc.org/10.1111/bjdp.12198

Sim, Z. L. & Xu, F. (2018). Another look at looking time: Surprise as rational statistical inference. *Topics in Cognitive Science, 11*, 154-163. https://doi.org/10.1111/tops.12393

Sobel, D. M., Benton, D., Finiasz, Z., Taylor, Y. & Weisberg, D. S. (2022). The influence of children's first action when learning causal structure from exploratory play. *Cognitive Development, 63,* 101194. https://doi.org/10.1016/j.cogdev.2022.101194

Sobel, D. M., Tenenbaum, J. B. & Gopnik, A. (2004). Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cognitive Science, 28*(3), 303-333. https://doi.org/10.1016/j.cogsci.2003.11.001

Spelke, E. S. (2024). Précis of what babies know. *Behavioral and Brain Sciences, 47,* e120. https://doi.org/10.1017/S0140525X23002443

Spelke, E. S. & Kinzler, K. D. (2007). Core knowledge. *Developmental Science, 10*(1), 89-96. https://doi.org/10.1111/j.1467-7687.2007.00569.x

Stahl, A. E. & Feigenson, L. (2017). Expectancy violations promote learning in young children. *Cognition, 163,* 1-14. http://dx.doi.org/10.1016/j.cognition.2017.02.008

Stahl, A. E., & Feigenson, L. (2019). Violations of core knowledge shape early learning. *Topics in cognitive science*, *11*(1), 136-153. https://doi.org/10.1111/tops.12389

Stahl, A. E. & Kibbe, M. M. (2022). Great expectations: The construct validity of the violation-of-expectation method for studying infant cognition. *Infant and Child Development, 31*(6). https://doi.org/10.1002/icd.2359

Tong, Y., Wang, F., & Danovitch, J. (2020). The role of epistemic and social characteristics in children's selective trust: Three meta-analyses. *Developmental Science, 23*(2). https://doi.org/10.1111/desc.12895

Tomasello, M. Carpenter, M. & Liszkowski, U. (2007). A new look at infant pointing. *Child Development, 78*(3). https://doi.org/10.1111/j.1467-8624.2007.01025.x

Török, G., Domberg, A. Ruggeri, A. (2025). Children adapt their information search to goal and statistical structure of a problem. *Developmental Psychology, 61*(9), 1607-1624. https://doi.org/10.1037/dev0001796

Ullman, T. & Tenenbaum, J. B. (2020). Bayesian models of conceptual development: Learning as building models of the world. *Annual Review of Developmental Psychology, 2*(1). https://hdl.handle.net/1721.1/138358.2

van Schijndel, T. J. P., Visser, I., van Bers, B. M. C. W. & Raijmakers, M. E. G. (2015). Preschoolers perform more informative experiments after observing theory-violating evidence. *Journal of Experimental Child Psychology*, *131*, 104–119. https://doi.org/10.1016/j.jecp.2014.11.008

Vriezen, E. R., & Pigott, S. E. (2002). The relationship between parental report on the BRIEF and performance-based measures of executive function in children with moderate to severe traumatic brain injury. *Child Neuropsychology*, *8*(4), 296-303. https://doi.org/10.1076/chin.8.4.296.13505

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

Weisberg, D. S. & Sobel, D. M. (2022). The evolution of rational constructivism. In D. S. Weisberg & D. M. Sobel (Eds.). *Constructing science: Connecting causal reasoning to scientific thinking in young children* (pp. 21-60). The MIT Press. https://doi.org/10.7551/mitpress/11939.001.0001

Whiten, A., McGuigan, N., Marshall-Pescini, S., & Hopper, L. M. (2009). Emulation, imitation, over-imitation and the scope of culture for child and chimpanzee. *Philosophical Transactions of the Royal Society B, 365*, 2417-2428. https://doi.org/10.1098/rstb.2009.0069

Wickham, H. (2016). Ggplot2: Elegant graphics for data analysis (R package version 4.5.1). Springer-Verlag. https://CRAN.R-project.org/package=ggplot2

Williamson R. A., Markman E. M. (2006). Precision of imitation as a function of preschoolers' understanding of the goal of the demonstration. *Developmental Psychology*, 42, 723–731. https://doi.org/10.1037/0012-1649.42.4.723

Willroth, E. G. & Atherton, O. E. (2024). Best laid plans: A guide to reporting preregistration deviations. *Advances in Methods and Practices in Psychological Science, 7*(1), 1-14. DOI: 10.1177/25152459231213802

Yu, Y., Landrum, A. R., Bonawitz, E., & Shafto, P. (2018). Questioning supports effective transmission of knowledge and increased exploratory learning in pre-kindergarten children. *Developmental Science, 21*(6): e12696. https://doi.org/10.1111/desc.12696

Zhu, H. (2024). *kableExtra: Construct complex table with kable and pipe syntax* (R package version 4.5.1). https://CRAN.R-project.org/package=kableExtra

Zsoldos, R. A. & Király, I. (2025). Pedagogy does not necessarily constrain exploration: Investigating preschoolers' information search during instructed exploration. *Developmental Science, 28,* e70004. https://doi.org/10.1111/desc.70004