Belief Revision in Children and Adults

By

Katherine Kimura

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor in Philosophy

in

Psychology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Alison Gopnik, Chair
Assistant Professor Celeste Kidd
Professor Fei Xu

Fall 2020

Belief Revision in Children and Adults

Abstract

Belief Revision in Children and Adults

by

Katherine Kimura

Doctor of Philosophy in Psychology

University of California, Berkeley

Professor Alison Gopnik, Chair

Learning is a complex process that requires integrating new evidence with existing beliefs. This task is relatively straightforward when the evidence is consistent with and thus supported by existing beliefs. In such instances the evidence serves to confirm and hence strengthen beliefs. It becomes less clear, however, when the evidence is inconsistent—that is, when the evidence directly conflicts with one's prior knowledge. Will learners revise their beliefs to reflect the new evidence or will they instead maintain their existing beliefs, perhaps ignoring the conflicting data as anomalous? How might their responses differ as a function of development, and what might this reveal about the mechanisms that support learning (and relearning)?

The current dissertation addresses these questions by asking how children and adults respond to evidence that violates their beliefs. Chapter 1 begins by describing a surprising developmental pattern: younger children are sometimes better than older children and adults at using statistical evidence to revise their beliefs. I then provide a Bayesian explanation for these findings, noting that younger learners may outperform older ones because they are intrinsically more flexible in their search for alternative hypotheses. Chapter 2 then provides empirical evidence that young children revise their higher-order beliefs by attending to the statistical strength of the data and weighing it against their prior belief. When shown a single event that conflicts with their belief, 4- and 5-year-olds largely maintain their higher-order belief. However, when shown several events that each support the same alternative hypothesis, children reliably update their belief to account for the new data. Chapter 3 then replicates these findings with even younger children, demonstrating that 3-year-olds similarly update their beliefs when the counterevidence is strong enough to outweigh the prior belief. Notably, children's ability to update their beliefs is independent of their executive ability to deliberately change their behaviors in response to new goals and demands. Chapter 4 then contrasts children's epistemic flexibility with adults' inflexibility, exposing a striking developmental decline in adults' sensitivity to new data. In particular, these findings highlight the robustness of adults' beliefs by demonstrating that adults continue to endorse a weakly supported belief even after they are primed or incentivized to

reason more broadly and creatively. Finally, Chapter 5 discusses the implications of this work and suggests future directions.

Taken together, these experiments contribute to our emerging understanding of belief revision by demonstrating how, as knowledge is accumulated and as beliefs are strengthened, our sensitivity to new evidence decreases and our willingness to revise declines. This pattern, although at times counterproductive, may be adaptive as learners transition from a strategy of exploration to one of exploitation.

To Paul, who lovingly steered the engine.

Table of Contents

List of Figures

List of Tables

Acknowledgements

First and foremost, I offer my sincerest gratitude to my advisor, Alison Gopnik. Her knowledge, which extends well beyond developmental science, is driven by a curiosity that is both inspiring and infectious. In addition to providing intellectual support, I am forever thankful for Alison's compassion during truly difficult times. Her patience has allowed me to build the confidence to ask deeper questions about who I am, what I believe, and how I want to live. I would also like to thank my other committee members, Fei Xu, Lara Buchak, and Celeste Kidd.

This work was possible by the support of various collaborators and labmates, including the many research assistants whose eagerness to engage with the research was a frequent reminder of its excitement. Our discussions, both about research and not, were as rewarding as they were fun.

I am also incredibly thankful for the friendships that were built on campus. My visits to Tolman Hall and to Berkeley Way West were always made a little sweeter (pun intended) after visiting Lisa Branum. Very early on Lisa became family, and ever since then she has been supporting me unconditionally and reminding me of what is truly important. Thanks must also go to Luvy Vanegas Grimaud, who was with me at the very end, cheering every time I collected a little more data and was that much closer to finishing. Thanks also to Frances Nkara. We journeyed through this program together, celebrating and commiserating with each other, and I am so grateful to have had Frances through it all. I must also thank Maria Baalbaki, who I first met at the Early Childhood Education Program while testing what is affectionately known as Doorbells. Throughout the years, Maria has continued to be my cheerleader, offering kindness, warmth, and compassion throughout this journey. I share that project with her.

This dissertation was possible because of my family both old and new. On numerous occasions, we created what felt like Santa's workshop to build toys and paint blocks for my studies. On even more occasions, we talked about my research and discussed its implications, which always brought it back in perspective and reminded me of its importance to the greater conversation. But, on most occasions, we didn't even talk about my work. Instead, we indulged in my newest obsessions and we celebrated our family expanding. They gave me a safe reprieve, a place to be me. This work is what it is because of the gentle support of my husband, Paul, who never once doubted me. His compassion for me and for others grounded me when I felt lost. I do not have the words, yet, to express how grateful I am for his love, care, and devotion over the years. This PhD is shared with him.

Thanks also to the families, teachers, and educators for volunteering their time and effort to this research. Lastly, this work was financially supported, in part, by the National Science Foundation Graduate Research Fellowship Program and by the Experimental Social Science Laboratory (Xlab) at UC Berkeley.

Chapter 1

Introduction

1.1 General introduction

Learning is a complex process that requires integrating new evidence with existing beliefs. This task is relatively straightforward when the evidence is consistent with and thus supported by existing beliefs. In such instances the evidence serves to confirm and hence strengthen beliefs. It becomes less clear, however, when the evidence is inconsistent—that is, when the evidence directly conflicts with one's prior knowledge (e.g., Bonawitz, van Schijndel, Friel, & Schulz, 2012). Will learners revise their beliefs to reflect the new evidence or will they instead maintain their existing beliefs, perhaps ignoring the conflicting data as anomalous? How might their responses differ as a function of development, and what might this reveal about the mechanisms that support learning (and relearning)?

It has been clear for a long time that children, including even young infants, are sensitive to evidence that directly violates their expectations about the world and contradicts their beliefs (e.g., Baillargeon, Spelke, & Wasserman, 1985; Wynn, 1992). In fact, previous studies show that such violations can actually promote learning by prompting selective exploration (Bonawitz et al., 2012; Stahl & Feigenson, 2015; van Schijndel, Visser, van Bers, & Raijmakers, 2015) and spontaneous explanation (Legare, Gelman, & Wellman, 2010). For example, 11-month-olds will systematically drop objects that seemingly violate their beliefs about support and meanwhile bang those that violate their beliefs about solidity (Stahl & Feigenson, 2015). Although several studies have since found that belief violations can actually enhance the learning of non-obvious properties (e.g., Stahl & Feigenson, 2015, 2017), only recently have studies demonstrated how belief violations then lead to systematic belief revision. It is not enough to simply detect (and be surprised by) inconsistent evidence. Learners must also generate alternative hypotheses that can explain their observations, evaluate these competing hypotheses against the data, and then decide how to update their beliefs given the interaction between the strength of the new evidence and the weight of their prior beliefs.

1.1.1 A surprising developmental decline

It is rather surprising that young children, despite having limited cognitive resources and poor executive function, are sometimes better than older children and adults at using new evidence to revise their beliefs (Lucas, Bridgers, Griffiths, & Gopnik, 2014). Consider, for example, a toy machine that plays music when particular combinations of blocks are placed on it. Preschoolers and adults initially thought that the individual blocks—and not the block combinations—caused the toy to activate. When shown, for instance, that the toy played music when two blocks were placed on top, both age groups tended to believe that one block was responsible and, when asked to activate the toy, selected a single block. If, however, children and adults were shown evidence to the contrary (i.e., that only specific combinations of blocks activated the toy, whereas individual blocks did not), children updated their former disjunctive belief to a conjunctive one and then generalized this newly revised belief to a set of new blocks. Now, when asked to activate the toy, they overwhelmingly selected two or more blocks. Adults, in contrast, continued to select a single block despite observing evidence that suggested otherwise. This striking decline in older learners' sensitivity to conjunctive evidence and the hardening of their disjunctive beliefs, although counterintuitive, is consistent across ages (Gopnik et al., 2017) and cultures (Wente et al., 2019).

This developmental pattern of younger learners outperforming older ones is also evident in other contexts with even younger ages. For example, 18-month-olds are more likely than 30-month-olds to learn abstract relational properties, such as *same* and *different* (Walker & Gopnik, 2014), and 4-year-olds are better than 6-year-olds at inferring social principles from patterns of evidence (Seiver, Gopnik, & Goodman, 2013). For instance, 4-year-olds rightfully predicted that a person behaved in a particular context (e.g., Sally avoided a diving board) because of a specific scenario (e.g., the diving board looked scary) and not because of a characteristic trait (e.g., Sally is scared). Six-year-olds, like adults, showed a strong bias for the trait explanation and failed to revise even when the evidence supported a situational explanation. Interestingly, children's sensitivity to social evidence, and hence their ability to update beliefs about social situations, reemerges during adolescence, a time in which they are particularly sensitive to social information (Gopnik et al., 2017). Similar to preschoolers, adolescents updated their social inferences based on data, whereas neither grade-schoolers nor adults did. Taken together, this developmental pattern, although consistent across several ages, domains, and contexts, is nonetheless surprising given that performance on cognitive tasks usually improves with age. Why in these cases do younger children reliably outperform older children and adults, and what does this pattern reveal about how these learning mechanisms develop?

1.1.2 A Bayesian explanation

Two possibilities may explain this developmental pattern, both of which are well described in terms of Bayesian learning. According to Bayes' rule, the probability that children and adults will revise their beliefs given some observed data depends on two factors: (1) their *prior probability*, or the degree to which they believe in a specific hypothesis before observing new data, and (2) the *likelihood*, or the probability that the observed data was generated by that particular hypothesis. Combining the prior probability with the likelihood yields the *posterior probability*, or the probability that an agent—child or adult—will revise her beliefs given the new data. This formalization leads to Bayes' rule:

$$P(h|d) = \frac{P(d|h)P(h)}{\sum_{h' \in H} P(d|h')P(h')}$$

in which $P(h)$ denotes the prior probability of hypothesis h, $P(d|h)$ indicates the likelihood that the observed data d was generated by h, and $P(h|d)$ represents the posterior probability. Thus, the probability that one will revise her beliefs is directly proportional to the product of the prior probability and the likelihood. In other words, as the probability of the prior distribution or likelihood decreases, so does the probability of the posterior distribution.

Since children have less knowledge than adults, they have flatter prior probability distributions and, therefore, have less reason to strongly believe in any one hypothesis. Thus, one possibility for why younger learners outperform older ones on certain tasks is that younger learners, because of their flatter priors, may require less evidence to overturn and revise their belief. If, for example, children hold a weak belief that individual blocks cause a toy to play music, then according to Bayes' rule children should have a low prior probability for the disjunctive rule, and therefore should also have a low posterior probability or belief that individual blocks activate the toy. As a result, when faced with enough evidence supporting the conjunctive or combination rule children should readily overturn their weak disjunctive belief. This, however, is in stark contrast with adults, who may have a higher prior probability distribution for the disjunctive rule because of their experience with causal mechanisms and, consequently, have a higher posterior probability. Adults may then require even more counterevidence to overturn their strong beliefs, whereas small amounts of data may suffice for children.

A second possibility for why children outperform adults is that children may be searching through their *hypothesis space H* (i.e., the range of all possible hypotheses) more widely than adults. With a wider search, children may be entertaining more hypotheses including those that are highly dissimilar to their prior. Adults, in contrast, may be only considering those hypotheses that are most consistent with their prior belief. In other words, children may be employing a strategy that favors wide search patterns and radical changes, whereas adults may by using a strategy that leads them to narrow searches and minor revisions. Although children's exploration may be beneficial when learning unusual hypotheses, it may be inefficient and perhaps even detrimental when only small adjustments are needed. It is possible then that the difference between children and adults' response to, for example, conjunctive evidence reflects a difference in their search patterns, with the conjunctive rule requiring a wider search than the disjunctive one.

It is important to note that these two possibilities are not mutually exclusive. In fact, those with stronger priors will be less likely to consider alternative beliefs and, therefore, less likely to search widely through their hypothesis space. In contrast, those with weaker priors will have flatter distributions and as a result will be more likely to search through their hypothesis space to consider alternative explanations. Although these possibilities are closely related, the ability to flexibly consider unusual alternatives may have an added benefit that is independent of prior knowledge.

1.1.3 Cognitive flexibility

The ability to consider alternative hypotheses seems to be inextricably linked to the prefrontal cortex. The prefrontal cortex, a region that is often associated with executive or cognitive control, is noted for biasing particular responses over others, thereby restricting the hypothesis space (Frith, 2000). As the prefrontal cortex matures, albeit very slowly, executive control improves and the ability to consider unusual, creative alternatives declines. It is worth noting that this top-down control is not necessarily an impediment to learning. In fact, as the prefrontal cortex matures and as cognitive control improves, learners develop strategies or heuristics that optimize the learning process and promote its efficiency. For instance, learners might take shortcuts by exploiting their past experience, using their prior belief to translate the present and make predictions about the future. Although these generalizations are useful when experiences are stable and the past can accurately inform the present, this efficiency can be costly when experiences are qualitatively different from one another. In these situations, generalizations using past experience may have the unintentional and counterproductive consequence of prematurely trapping learners, preventing them from making new discoveries (Rich & Gureckis, 2018). In such cases, it may be helpful to learn without the constraints imposed by the prefrontal cortex.

It appears that young children are particularly adept at avoiding these learning traps, perhaps because of their immature prefrontal cortices (Thompson-Schill, Ramscar, & Chriskiou, 2009). Since the prefrontal cortex is late to mature, young children show considerable deficits in their executive control. At the same time, however, they show remarkable creativity when problem solving. Indeed, preschoolers are actually better than older children and adults at solving classic insight problems, whose unconventional solutions often require some degree of ingenuity (German & Defeyter, 2000). When asked, for instance, to reach for a toy on a high shelf, 5-year-olds creatively repurpose a box, overturning it to create a stool. Six- and 7-year-olds, in contrast, have difficulty producing an alternative function for the box, fixating instead on its conventional use. This flexibility by way of reduced executive control is also apparent in children's selective attention. Unlike adults who are better at attending to goal-directed information, preschoolers are better at processing information that is seemingly irrelevant (Plebanek & Sloutsky, 2017).

This privileged relation between the prefrontal cortex and cognitive flexibility is even evident in adults with limited cortical activity. Adults with damage to their lateral prefrontal cortex are, like children with immature cortices, surprisingly better than matched controls at solving insight problems (Reverberi, Toraldo, D'Agostini, & Skrap, 2005). In particular, when adults are presented with a false arithmetic statement constructed out of sticks and displayed as Roman numerals (e.g., III = III + III), those with damage to their prefrontal cortex are better at moving a single stick from one position to another to transform the false statement into a valid one (e.g., III = III = III). Similar advantages to problem solving have been demonstrated when cortical activity is suppressed in healthy adults with non-invasive transcranial electrical currents (Chi & Snyder, 2011, 2012; Chrysikou, Hamilton, Coslett, Datta, Bikson, & Thompson-Schill, 2013; Weinberger, Green, & Chrysikou, 2017).

Although limited activity in the prefrontal cortex—by virtue of being young, by injury, or by suppressed stimulation—can promote flexibility and enhance problem solving, there is considerable research demonstrating that children, especially those with poor executive control, are actually cognitively inflexible (e.g., Carlson, 2005; Diamond, 2013; Garon, Baryson, & Smith, 2008; Zelazo et al., 2003). For example, when asked to sort bivariate cards first by one rule (e.g., color) and then by another (e.g., shape), 3-year-olds flawless sort by the first rule but, when asked to switch, they perseverate and fail to sort by the other rule. Older children, however, can change their sorting behavior when instructed, although even they have difficulty when the rules become more complex (Zelazo, 2006). This apparent paradox begs the question as to why children with limited executive control appear cognitively flexible in one context but cognitively inflexible in another. To answer this question, I distinguish between two types of flexibility—epistemic flexibility and executive flexibility—arguing that these two are independent of one another.

In this dissertation, I examine how children and adults respond to new evidence, contrasting children's remarkable flexibility with adults' remarkable rigidity. In doing so, I demonstrate how learning strategies shift across development as the prefrontal cortex matures, flexibility narrows, and cognitive control improves. The delicate balance between these two strategies helps children and adults optimize learning across the lifespan, allowing children to easily learn new information while stabilizing adults' beliefs to ease inferences and predictions. The chapters that follow focus on the mechanisms that we, as both children and adults, use to learn and relearn new knowledge.

1.2 Précis

This dissertation investigates how children and adults respond to evidence that directly violates their beliefs, focusing on the factors that encourage as well as discourage belief revision across development.

Chapter 2 begins to examine this complex learning process by exploring whether children can revise their higher-order beliefs when shown evidence that conflicts with their existing beliefs and, furthermore, whether they do so by attending to both the weight of the evidence and the strength of their beliefs. Consistent with Bayesian learning, we find that 4- and 5-year-olds are rational revisionists. When shown a single event that conflicted with their prior beliefs, children largely maintained their initial higher-order belief. However, when shown several events that each supported the same alternative hypothesis, children systematically updated their beliefs to account for the new data. These findings suggest that children not only revise their higher-order beliefs but do so rationally.

Although young children are rational, flexible thinkers, they are also, in other contexts, stubbornly inflexible (e.g., Zelazo et al., 2003). When instructed to sort first by one rule and then by another, they often perseverate and fail to switch between rules and tasks. Chapter 3 explores this paradox by distinguishing between two types of cognitive flexibility: (1) *epistemic flexibility*—the ability to change beliefs in response to new information—and (2) *executive flexibility*—the goal-directed, top-down ability to shift to new demands. Across three experiments, we ask whether 3-year-olds, who typically have difficulty switching between rules because of their limited executive function (Diamond, 2013, Garon et al., 2008), can revise their beliefs in response to new data. Consistent with Chapter 2, we show that even 3-year-olds can flexibly and rationally change their beliefs when presented sufficient counterevidence. We further report that this normative revision is independent of their emerging executive function, suggesting that there may be two distinct types of cognitive flexibility: the epistemic ability to flexibly update beliefs in response to new data, and the executive ability to use top-down control processes to intentionally override goal-directed behaviors.

Whereas Chapters 2 and 3 highlight the rationality of children's beliefs, demonstrating that young children change their minds on the basis of the evidence, Chapter 4 exposes a striking developmental shift in the ability (or inability) to revise beliefs by highlighting adults' unwillingness to revise in response to new data. Across three experiments, we demonstrate the apparent difficulty in encouraging flexible, accurate belief revision in adults. For instance, we asked adults to generate either 1 or 6 descriptions for common objects with the assumption that generating multiple descriptions would require a wider search than generating one. Afterwards, participants observed a novel toy activate by an unusual rule (i.e., specific combinations of blocks caused a toy machine to play music, whereas individual blocks did not). Despite observing evidence supporting the more unusual belief that pairs of blocks activated the toy (i.e., conjunctive rule) and despite being primed to reason more broadly, adults failed to revise their more commonly held belief that a single block was causally responsible (i.e., disjunctive rule). We also found that adults continued to endorse this disjunctive rule even after they observed conjunctive evidence across multiple domains and even after they were incentivized to respond correctly. These results, which are in stark contrast to the normative flexibility of children's beliefs, demonstrate the remarkable rigidity as well as robustness of adults' beliefs.

Together, these studies illustrate how, as learners mature, their sensitivity to new evidence and their flexibility to revise beliefs decline. Although children respond rationally to information that violates their beliefs, adults seemingly do not. This striking developmental pattern highlights the difference between children and adults as experience strengthens prior beliefs and as learners transition from a strategy of exploration to one of exploitation.

Chapter 2

Rational higher-order belief revision in young children

2.1 Introduction

Children are remarkable learners. Despite their limited knowledge, children can form higher-order beliefs, or overhypotheses (Goodman, 1955), about the world from surprisingly few examples (e.g., Dewar & Xu, 2010; Kemp, Perfors, & Tenenbaum, 2007; Lucas, Bridgers, Griffiths, & Gopnik, 2014; Sim & Xu, 2017). For instance, children might come to believe that all dogs bark, all cows moo, and all pigs oink, while also forming the higher-order belief that every animal species makes a unique noise that is different from other animal species. This higher-order belief then constrains specific lower-level beliefs. When presented with an unfamiliar animal, children could predict that this animal will produce a sound that is uniquely distinct from barking, mooing, and oinking. Suppose, however, that children observe evidence that violates this prediction (e.g., a barking sea lion). How will they update their higher-order belief to deal with this new information?

Children, including young infants, are sensitive to evidence that directly conflicts with their beliefs. In fact, previous studies suggest that belief violations can enhance learning by encouraging spontaneous explanations (Legare, Gelman, & Wellman, 2010; Legare, Schult, Impola, & Souza, 2016) and selective exploration (Bonawitz, van Schijndel, Friel, & Schulz, 2012; Stahl & Feigenson, 2015; van Schijndel, Visser, van Bers, & Raijmakers, 2015). Even in the absence of explanation and exploration, children can revise their beliefs in response to new evidence. For example, preschoolers can use belief-violating evidence to update their understanding of balance (Bonawitz et al., 2012b), theory of mind (Amsterlaw & Wellman, 2006; Slaughter & Gopnik, 1996), and the logical form of causal relations (Lucas et al., 2014).

Children also appear to normatively revise their beliefs in ways that are consistent with Bayesian learning (Kushnir & Gopnik, 2007). When four-year-olds are asked to activate a novel toy with a block, they overwhelmingly place the block directly on top of the toy. However, after observing that a block held above the toy was *more* likely to activate the toy than a block placed directly on top, these children revised their prior belief. When asked to activate the toy again, this time with a new block, they appropriately held the object over the toy.

Similarly, Schulz, Bonawitz, and Griffiths (2007) showed that young children will normatively revise their lower-level beliefs about causal relations that cross domains. Preschoolers initially deny that psychological factors (e.g., worry) can result in physical responses (e.g., stomachache), expecting instead that physical outcomes are the result of physical causes (e.g., eating bad food; Notaro, Gelman, & Zimmerman, 2001). If, however, children are shown multiple instances supporting a psychosomatic cause, then four- and five-year-olds will revise their prior belief and endorse this cross-domain causal relation. This learning, however, is relatively conservative, as children will often fail to generalize their revised belief to a new psychosomatic event (e.g., nervousness causes sickness). That is, revising their beliefs about one causal relation does not increase their willingness to accept other similar relations.

Although it is clear that children can revise their beliefs in response to counterevidence, there are at least two aspects of belief revision that remain unclear. First, it is unclear whether children are revising their higher-order beliefs (e.g., psychological factors can cause physical symptoms) or are instead simply updating their lower-level beliefs (e.g., worry can cause stomachaches). If children are revising their higher-order beliefs, then they should be able to apply their updated belief to other lower-level examples provided that these are consistent with and thus supported by their higher-order belief. Alternatively, children might be simply updating their lower-level beliefs, thereby inferring that the observed evidence applies narrowly in scope. Thus, one goal of the current study is to examine whether children can actually update their higher-order beliefs given conflicting information or whether they will only update their lower-level beliefs.

Second, it is unclear how conflicting evidence interacts with prior beliefs to promote higher-order belief revision. Bayesian learning suggests that the probability that learners will revise their higher-order belief depends on both the prior probability of that belief and the strength of the evidence. If the prior probability is high, then more evidence may be required to overturn the belief. But if the prior probability is low, then small amounts of data should suffice. Notably, however, earlier studies of belief revision examined learners' response to evidence that violates their current theories, such as their beliefs about causal spatial contiguity or psychosomatic causes. These beliefs will vary in strength depending on the individuals' experience, making it unclear whether the decision to revise reflects a normative Bayesian interaction between prior beliefs and observed evidence.

Indeed, children's decisions about whether to revise might be independent of the strength of their prior beliefs and the new evidence. For instance, children might be irrationally inflexible in their belief revision by requiring large amounts of data to change their beliefs (e.g., Zelazo, Frye, & Rapus, 1996). Alternatively, they might be irrationally willing to revise, changing their beliefs regardless of the strength of the evidence. It is also possible, however, that children might show a rational balance between the strength of their prior beliefs and the strength of the new evidence. In the current study, we explore these questions by systematically controlling the evidence children observe for an initial higher-order belief and the evidence that then contradicts that belief.

We investigate whether children can revise their higher-order beliefs and, furthermore, the extent to which children will revise their beliefs by attending to both the weight of their prior and the causal strength of the evidence. Children were shown pairs of novel toys and were taught a deterministic rule that supported a higher-order belief about how the toys worked. Participants then observed evidence supporting a different deterministic rule that conflicted with the first. Finally, they were asked to activate a novel toy by selecting from among three blocks, which either supported or failed to support either rule. If children revise their higher-order beliefs as a function of the both their priors and the evidence, then they should revise their higher-order beliefs when the conflicting evidence is stronger than the initial evidence and, likewise, maintain their higher-order beliefs when the initial evidence is stronger than the counterevidence. If, however, they are not sensitive to the strength of the evidence, then they should be just as likely to update (or maintain) their higher-order beliefs in both these cases.

2.2 Method

2.2.1 Participants

Ninety-six preschoolers (*M* = 4;10, range = 4;0–5;8, 42 females) from the San Francisco Bay Area participated in this study between April 2015 and May 2017, including fifty-seven 4-year-olds (*M* = 4;6, range = 4;0-4;11, 25 females) and thirty-nine 5-year-olds (*M* = 5;4, range = 5;0-5;8, 17 females). Children were randomly assigned to either a condition with one learning trial and four revision trials (*n* = 27 4-year-olds, 11 females; *n* = 23 5-year-olds, 9 females) or to a condition with four learning trials and one revision trial (*n* = 30 4-year-olds, 14 females; *n* = 16 5-year-olds, 8 females). Children were from predominantly middle- to upper-class families and were recruited from and tested at local preschools. An additional 15 children participated but were excluded from the analysis for experimenter error (*n* = 5), for failure to complete the task (*n* = 4), or for accidental malfunction of the toys (*n* = 6).

2.2.2 Materials

Sixteen toys (6” x 6” x 1.5”), each equipped with a hidden wireless doorbell, were organized into eight identical pairs. Modeled after the blicket detector paradigm (Gopnik & Sobel, 2000) with procedures similar to those used by Sim and Xu (2017), it appeared that certain blocks ‘activated’ the toys, making them play music. In fact, the toys were surreptitiously activated by the experimenter.

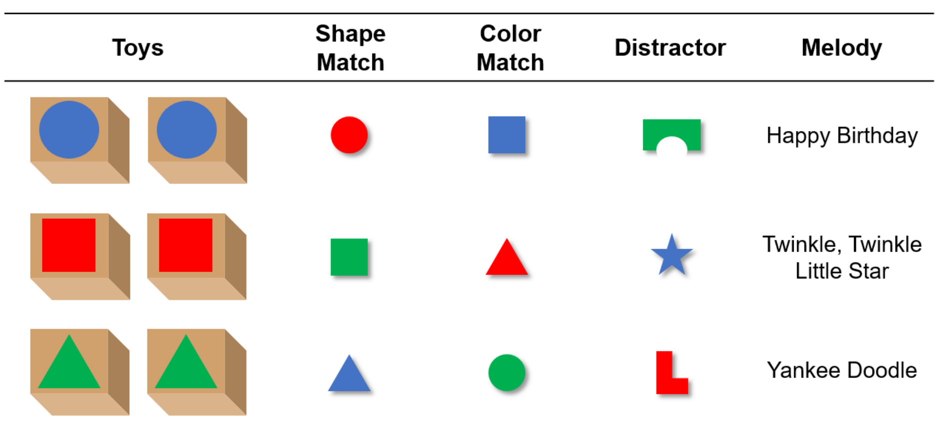

Figure 2.1 Sample stimulus sets.

Glued to the top of each toy was a flat wooden piece (4.5” x 4.5”). These pieces varied across pairings such that each pair had a unique color and shape that was different from the other pairs. For example, one pair had *blue circles*, whereas another pair had *red squares*. The toys played a different melody depending on the piece on top. For instance, the toys with blue circles played *Happy Birthday* when activated, whereas the toys with red squares played *Twinkle, Twinkle Little Star* (see Figure 2.1).

Three wooden blocks (1.5” x 1.5” x 1.5”) corresponded to each pair of toys: (1) a shape match, (2) a color match, and (3) a distractor. One of the blocks—the shape match—matched the piece on top of the toy in shape but not in color (e.g., a red *circle* block matched the blue *circle* toy), whereas the other block—the color match—matched in color but not in shape (e.g., a *blue* square block matched the *blue* circle toy). The third block—the distractor—did not match the toy in either shape or color (e.g., a green bridge block did not match the blue circle toy in either shape or color). Critically, none of the blocks matched any of the toys on both shape and color.

2.2.3 Design

Participants were randomly assigned to one of two conditions, which differed only in the number of training trials administered during the learning and revision phases. One condition consisted of one training trial in the learning phase and four in the revision phase. In the other condition, the learning phase had four training trials, whereas the revision phase had only one. The activation rule in each phase (color vs. shape) was counterbalanced across participants. Thus, the experimental design was 2 (training trials: 1 vs. 4) x 2 (rule: color vs. shape) with the number of training trials in each phase and the activation rule as between-subjects variables.

2.2.4 Procedure

Children were seated at a table directly across from the experimenter. Each session lasted approximately 15 minutes and consisted of two phases: (1) a learning phase and (2) a revision phase. In the learning phase, children saw evidence that supported a particular higher-order belief about how the toys worked. This established a common set of priors for all the children. In the revision phase, they were shown evidence supporting a different higher-order belief that conflicted with the previous one. For example, children might initially learn that the toys activate when the blocks match the toys' top piece in shape but not in color; however, in the revision phase, children would observe the toys activate with blocks matching in color rather than shape.

Across the two phases, children selected seven pairs of toys: five were observed as training trials, whereas two were used for test trials. Once a pair of toys was selected and the trial was completed, that pairing was removed from the selection for subsequent trials so that a new pair of toys was used for each trial. Both the learning and revision phases began with either one or four training trials and ended with a single test trial.

2.2.4.1 Learning phase

The experimenter began the learning phase by introducing children to her toys: "Look, these are my toys! Sometimes when I put blocks on top of my toys they play music. And sometimes when I put blocks on top of my toys they do not play music." The experimenter then invited the child to play her game and figure out how the toys worked.

For each training trial of the learning phase, the experimenter presented photographs, one for each pair of toys, and asked the child to select a pair by pointing to its picture. By asking children to select the toys and, hence, decide the presentation order, we discouraged the inference that the experimenter had already separated toys that worked by one rule from those that worked by another. After the child responded, the experimenter placed the selected toys on the table and placed two corresponding blocks—the shape match and the color match—directly in front of the first toy, randomizing the left-right placement across trials. The experimenter then, one at a time, placed each block on top of the toy, which activated according to either a shape rule (e.g., *circle* blocks activate *circle* toys) or a color rule (e.g., *blue* blocks activate *blue* toys). To emphasize the outcome, the experimenter would smile and say, “It turned on! It’s playing music!” when the toy activated, whereas she would shake her head and say, “No music. It didn’t turn on.” when it did not activate. After the effect of both blocks was demonstrated—one which resulted in activation and one that did not—the experimenter moved the first toy aside and placed the same blocks on the second toy, again doing so one at a time. This toy, which was identical to the first one, activated according to the same rule. The experimenter then removed these toys and, for children who observed four training trials, repeated the procedure with three new pairs of toys.

After completing the training trials, the experimenter administered a forced-choice test trial to assess whether or not the child learned the causal rule and, critically, generalized this rule to other novel toys. Similar to the training trials, the test trial began with the experimenter asking the child to select a new pair of toys by pointing to its picture. However, unlike the training trials, the experimenter presented only one of the two toys along with three test blocks—the shape match, the color match, and the distractor. These blocks were arranged on a tray in a random presentation order in front of the toy. The experimenter then asked the child to demonstrate his or her higher-order belief about the toys by asking, “Can you point to what will make my toy play music?” as she moved the tray towards the child. After the child responded, the experimenter provided a neutral response, removed the toy and blocks, and proceeded to the revision phase.

2.2.4.2 Revision phase

The revision phase was identical to the learning phase with the following exception: the training trials in the revision phase supported the opposite rule to that demonstrated in the learning phase. For instance, if children initially learned that blocks of the same color activated a set of toys, then in the revision phase they observed that blocks of the same shape activated the remaining set of toys, while blocks of the same color did not. As in the learning phase, a final forced-choice test trial assessed whether participants revised or maintained their original higher-order belief.

2.2.5 Coding

Responses were coded based on whether the child selected the block corresponding with the observed rule for that given phase. For example, children who observed the color rule and selected the color match in the learning phase were coded as learning the higher-order rule, whereas children who selected the shape match or the distractor after observing identical evidence were coded as not. In the revision phase, responses were similarly coded as being consistent or inconsistent with the revision rule. The first author performed the coding, whereas a second coder blind to the condition and activation rule scored a random subset of participants (40%). Interrater reliability was 100%.

2.3 Results

Preliminary analyses revealed no significant difference between children who observed the color rule and selected the color match from those who observed the shape rule and selected the shape match, thus implying that children were just as likely to learn the color rule as they were the shape rule. There were also no significant differences in performance between four- and five-year-olds. The data were, therefore, collapsed across learning rules and across age groups. All analyses were two-tailed.

2.3.1 Learning phase

Despite this markedly challenging task, especially for participants who observed a single training trial, children selected the block corresponding with the learning rule more often than would be expected by chance (i.e., .33) after observing one training trial (30 out of 50 children; 60.0%) $p < .001$, binomial test), or four training trials (32 out of 46 children; 69.6%), $p < .001$, binomial test. Although children who observed four training trials observed more evidence in support of the learning rule, the difference between the two conditions was not significant, $p = .395$, Fisher's exact test.

Although this consistent and correct responding suggests that children inferred the appropriate causal relation, learning a higher-order rule from few observations was, nevertheless, a challenging task. When they saw only one training trial, 20 children out of 50 (40.0%) failed to infer the correct rule and, even when they saw four consecutive training trials that supported the same higher-order rule, 14 children out of 46 (30.4%) still failed to learn the rule.

2.3.2 Revision phase

We next examined whether or not children revised their causal beliefs given the strength of the conflicting evidence. To do this, we restricted analyses to include only those children who learned the initial rule (one learning trial condition: $n = 30$; four learning trials condition: $n = 32$) and tested whether they varied their revision patterns as a function of the number of training trials (i.e., 1 vs. 4 trials). If participants were weighing the strength of the evidence against the strength of their prior beliefs, then they should be *more* likely to revise their beliefs when shown one learning and four revision trials than if shown four learning and one revision trial.

Consistent with these predictions, we found that children who observed one learning trial and four revision trials were more likely to select the block that corresponded with the revision rule (19 out of 30 children; 63.3%) than those who observed four learning trials and one revision trial (7 out of 32 children; 21.9%), $p = .002$, Fisher's exact test. Furthermore, when presented with one learning and four revision trials, children selected the block corresponding with the revision rule at a rate that exceeded chance performance, $p < .001$, binomial test, indicating that they reliably revised their higher-order beliefs. In contrast, children who observed four learning trials and one revision trial were more likely to select the block that was consistent with the learning rule (22 out of 32 children; 68.8%), $p < .001$, binomial test. In other words, when the prior was weaker and the counterevidence was stronger, children were likely to revise their beliefs appropriately. However, when the prior was stronger and the counterevidence was weaker, children largely maintained their initial higher-order beliefs.

2.4 Discussion

These analyses indicate that children can revise their higher-order beliefs in a way that is consistent with Bayesian learning. When children observed pairs of novel toys that activated according to a particular rule, thereby supporting a higher-order belief, they were more likely to maintain this belief if, overall, they viewed more evidence that supported their initial rule than the revised rule. However, if they instead observed more evidence that supported the conflicting rule, then children would systematically revise their belief to appropriately account for the new data.

Similar to previous studies (e.g., Bonawitz et al., 2012b; Kushnir & Gopnik, 2007; Legare et al., 2016; Lucas et al., 2014; Schulz et al., 2007), we found that children can quickly revise their beliefs in response to conflicting evidence. However, unlike previous studies, we explicitly demonstrate that children can revise their higher-order beliefs. For example, observing the blue block activate the blue toys prompts children to form a higher-order belief—e.g., blocks that correspond in color will activate the toys. This higher-order belief then constrains and guides their beliefs about novel toys in the test trial. For instance, children will predict that the red block will activate the red square toy rather than the square block. If, however, they observe that the red block fails, whereas the square block succeeds—and they also observe other instances that support the alternative shape rule—then children will update their higher-order belief. On the test trial, they will now predict that blocks corresponding in shape, and not in color, will activate the toys.

2.4.1 Limitations and future directions

Although our findings suggest that children revised their higher-order beliefs, we do not know whether they were also revising their lower-level beliefs. Children might, for instance, continue to rightfully believe that blue blocks activate blue circle toys even though this lower-level belief conflicts with their newly revised higher-order belief. It is also possible that children are updating both their lower- and higher-order beliefs. In addition to now believing that blocks matching in shape activate the toys, children might, for instance, believe that circle blocks activate blue circle toys. In fact, lower-level belief revision might precede higher-order belief revision, particularly if higher-order beliefs are more stable and less susceptible to change. Alternatively, higher-order beliefs, which often have few competing alternatives (Perfors, Tenenbaum, Griffiths, & Xu, 2011), could be revised first. These questions are important in deepening our understanding about how observed evidence interacts with complementary and competing beliefs but answering them will depend on further research.

We have focused on how learners respond to evidence that behaves deterministically; however, the evidence we typically encounter behaves stochastically. How might learners reason about such probabilistic evidence when revising their higher-order beliefs? Our experiment does not directly address this question, but it seems plausible that the strategies learners employ when they evaluate deterministic evidence are similar to those they use when they weigh probabilistic evidence. Indeed, children might have reasoned that the toys behaved probabilistically when they combined the evidence from both the learning and revision phases. In fact, children might think that the system instantiated a higher-order regularity and that there were occasional exceptions to that regularity—i.e., the one conflicting trial might be a fluke (Nosofsky, Palmeri, & McKinley, 1994). As a result, children would maintain their initial belief when the evidence strongly favored that belief—even when there is an occasional exception.

Although we found no difference between four- and five-year-olds in their ability to revise their higher-order beliefs, many children failed to initially learn the higher-order belief and, therefore, could not be included in the revision phase. It is thus possible that the children we tested—i.e., those who learned the higher-order rule—were particularly advanced compared to four- and five-year-olds in general. There is also evidence suggesting that younger children, three-year-olds in particular, might have more difficulty revising their beliefs in response to new data. In Kushnir & Gopnik (2007), three-year-olds, like four-year-olds, initially believed that causation required direct contact. Although the older children eventually revised this belief when they saw counterevidence, the younger children had difficulty overriding their prior preference, even after they saw deterministic, unambiguous evidence to the contrary. Similarly, three-year-olds, in contrast to older children, failed to endorse psychosomatic causes and instead continued to demonstrate a strong bias towards the domain-appropriate cause. Younger children only revised their beliefs after a two-week training, during which they were taught to reason about statistical evidence (Bonawitz, Fisher, & Schulz, 2012). This is consistent with other evidence suggesting that three-year-olds are relatively inflexible. For instance, when preschoolers are instructed to sort a series of bi-dimensional cards first according to one dimension and then according to another, they often perseverate and fail to switch to the second dimension (Zelazo et al., 1996). In future studies, we plan to extend this work to younger children to explore whether there is a qualitative difference between three- and four-year-olds, as suggested by the earlier studies, and whether this difference is related to performance on dimensional shift tasks. It would also be interesting to explore other potential correlates of this kind of flexibility, such as bilingualism, culture, socioeconomic status, or executive control.

2.4.2 Conclusion

Belief revision is a highly complex and dynamic process that involves evaluating the evidence against prior beliefs at multiple levels of abstraction. Despite these challenges, our findings suggest that children can revise their higher-level beliefs when they see conflicting evidence and, indeed, do so rationally. When the evidence strongly supports an alternative, children overwhelmingly revise their beliefs to reflect the new information. These higher-order beliefs, however, are relatively robust. When there is more evidence for the initial belief than for the alternative, children maintain their initial belief. Although our knowledge is relatively stable, every once in a while, when counterevidence evidence is strong enough, it can lead to a new discovery that radically alters our thinking.

Chapter 3

Epistemic vs. executive flexibility: Evidence for separate forms of flexibility in belief revision and executive control

3.1 Introduction

Young children are seemingly flexible, changing their minds in response to new observations and, rather remarkably, doing so rationally (e.g., Kimura & Gopnik, 2019). Yet, at the same time, they are seemingly inflexible. When situational demands change, they are unable to switch to new tasks and perspectives (e.g., Zelazo et al., 2003). Why do children appear cognitively flexible in one context yet cognitively inflexible in another? In this paper, we distinguish between epistemic flexibility—the ability to change beliefs in response to new information—and executive flexibility—the goal-directed top-down ability to shift to new demands. We begin with a brief discussion about epistemic flexibility, noting that children normatively update their beliefs in ways that are consistent with Bayesian learning when their expectations are violated. We then contrast these findings with a review of executive function and describe how young children often repeat behaviors that are no longer appropriate. This distinction leads to our hypothesis, which we go on to test in three experiments, that these are two distinct forms of cognitive flexibility. Both forms can lead to changes in behavior, but they are, nonetheless, independent of one another.

Research suggests that children apply Bayesian statistical inference to acquire the meanings of new words (Xu & Tenenbaum, 2007), to infer causal relations (Kushnir & Gopnik, 2007; Lucas, Bridgers, Griffiths, & Gopnik, 2014; Schulz, Bonawitz, & Griffiths, 2007), and to form abstract inductive constraints to guide generalizations (Dewar & Xu, 2010). This kind of Bayesian learning involves a balance between old knowledge and new evidence, and hence between stability and flexibility. Bayesian inference allows a learner to rationally determine when to stick with their current beliefs and when to revise them in the face of new evidence. This kind of inductive balance applies throughout the learning process, from the initial formation of knowledge to its continued revision as new information is integrated with the old.

Indeed, recent findings reveal that children are rational when they evaluate new, contradictory evidence. When young children see a small amount of evidence that violates their current belief, they maintain that belief, for example, by explaining away the counterevidence (Bonawitz, van Schijndel, Friel, & Schulz, 2012; Schulz, Goodman, Tenenbaum, & Jenkins, 2008) or by interpreting the data as anomalous (Kushnir, Xu, & Wellman, 2010). Preschoolers can, however, change their minds to account for the new data if they observe enough counterevidence to outweigh the prior belief (Kimura & Gopnik, 2019; Schulz et al., 2007). Even younger children, particularly 3-year-olds who seem resistant to change, can update their beliefs in response to strong statistical evidence that is taught over multiple training sessions (Bonawitz, Fischer, & Schulz, 2012; Rhodes & Wellman, 2013). Whereas small amounts of evidence may be sufficient to overcome weak beliefs, larger amounts may be necessary to revise stronger ones.

A recent paper shows that children are remarkably sensitive to the precise strength of the evidence when they rationally update their higher-order beliefs (Kimura & Gopnik, 2019). When four- and five-year-olds were shown, for example, that a red block, placed on top of a red square toy, makes the toy play music, they quickly formed the higher-order belief that the toy activated when the color of the toy matched the color of the block. Afterwards, when they were asked to activate a new toy—one they had never seen before—children generalized this newly formed belief and selected the corresponding color block (e.g., selecting a blue block to activate a blue circle toy). Then children observed a series of new toys that behaved contrary to this belief. Instead of observing the toys activate with their matching color blocks, children saw that the toys activate when the shape of the toys matched the shape of the blocks. Critically, children were never told about activation rules nor was the rule switch ever highlighted. If, overall, children saw more evidence that the new toys followed the shape rule than the color rule, they would systematically update their belief to account for the new data. Now, when they were asked to activate a different toy, they selected the corresponding shape block. If, however, children had observed more evidence that supported the initial color rule than the conflicting shape rule, then they would maintain their original higher-order belief and continue selecting the color block. This suggests that children rationally updated abstract higher-order beliefs by balancing the statistical strength of the new evidence with the weight of the prior belief.

Despite being epistemically flexible, young children have difficulty changing perspectives and shifting behaviors in response to new top-down demands (Zelazo et al., 2003). This inflexibility has been largely associated with children's immature executive functions, and has been described as a deficit in "cognitive flexibility" (Diamond, 2013; Garon, Bryson, & Smith, 2008). This ability to deliberately switch between arbitrary rules and tasks improves dramatically during the preschool years. For example, in the standard Dimensional Change Card Sort (DCCS) task, 3- and 4-year-olds are asked to sort cards first according to one rule (e.g., color) and then according to another (e.g., shape). Three-year-olds can flawlessly sort by the first rule but, when they are asked to switch and sort by the second rule, most perseverate and continue to sort by the first rule. Four-year-olds, in contrast, can begin to flexibly switch between the two rules when instructed to do so, demonstrating their increased ability to inhibit irrelevant dimensions, maintain representations of the rules in their working memory, and switch between them (Zelazo et al., 2003), although even four-year-olds struggle when the rules become more complex (Zelazo, 2006). Similar age-related improvements have been reported across a suite of executive function tasks, including children's ability to reverse perspectives and to suppress prepotent responses (see Carlson, 2005 for a review).

At least superficially, the Kimura and Gopnik belief revision task and the DCCS task appear to be similar. In both cases, children begin by sorting objects based on one higher-order rule, shape or color, and then must switch to the opposite rule, either because they have seen countervailing evidence or because they are instructed to do so. What is the relation between these two types of tasks? It is possible that performance on both tasks reflects a general flexibility that increases with age. Kimura and Gopnik found that 4- and 5-year-olds spontaneously changed their beliefs when presented with new evidence. This parallels the generally successful performance of these older preschoolers in the DCCS task, in contrast to 3-year olds. If age-related improvements explain performance on both types of tasks, then 3-year-olds should fail to revise their beliefs despite strong counterevidence and their performance on the belief revision and executive function tasks should improve with age and should be correlated.

Another possibility is that performance on the DCCS is negatively associated with some kinds of children's belief revision. In particular, executive function may narrow the hypothesis search by filtering out competing alternatives, explanations, and beliefs (Thompson-Schill, Ramscar, & Chrysikou, 2009). Learners with poor executive function may then have weaker constraints when interpreting new data and, as a result, may generate more alternative hypotheses to explain conflicting evidence such as those that are unusual or uncommon. Indeed, younger children, who have less executive function than older ones, are better than older children and adults at solving insight problems that require creative solutions (German & Defeyter, 2000) and at inferring unusual causal relations from data (Lucas et al., 2014). If children are flexible learners precisely because of their limited executive function, then we might suspect that 3-year-olds with limited executive function should have greater epistemic flexibility.

It is also possible, however, that the belief revision task and the DCCS task measure two distinct types of flexibility. On the one hand, epistemic flexibility captures changes in knowledge such as beliefs. Consider, for example, a child with the belief that the toy follows the color rule. Suppose then that this child observes evidence that routinely contradicts this belief. In order to resolve this apparent conflict between the new evidence and the current belief, the child must first detect the inconsistency and notice that, contrary to expectation, the new toy consistently fails to play music with matching color blocks but succeeds with matching shape blocks. How the child responds to this evidence depends on two distinct abilities. First, the child must generate an alternative hypothesis—e.g. toys activate with their matching shape block. Second, the child must evaluate this alternative against the prior belief—e.g., more toys activate with their shape blocks than with their color blocks—perhaps by applying Bayesian learning.

On the other hand, executive function refers to the top-down processes that are recruited when deliberately controlling (and shifting) attention, perspectives, and behaviors. When instructed to sort bivariant cards by the color rule, for instance, the child holds in their working memory a representation of the rule and can seamlessly place red cards in one tray and blue cards in the other. But, when they are asked to switch their behavior and sort those same cards by the shape rule, the children must inhibit their attention to the color representation, hold a new shape representation in their working memory, and make the appropriate behavioral switch. Beliefs about the rules do not change. Rather the experimenter arbitrarily changes the external demands and the child's goal-directed behaviors change as a result. Children who perseverate on this task may be inflexible because they have difficulty inhibiting, holding, or manipulating these representations even if they are able to change these representations in the light of new evidence (Kirkham, Cruess, & Diamond, 2003; Kloo & Perner, 2005; Zelazo et al., 2003).

Other evidence suggests that the ability to change beliefs in response to new evidence and the ability to change behaviors in response to new demands may very well be independent of one another. Belief revision involves tracking statistical regularities in the environment to draw inferences about the data. Even infants can make sophisticated inferences from statistics (for a review, see Saffran & Kirkham, 2018). They can parse speech into words based on the probabilistic frequency of adjacent syllables, and they can infer the probability of a population from the probability of a sample. Yet infants' executive functions are rudimentary at best (for a review, see Garon et al., 2008). Given that the onset of these two processes vary considerably, we might expect that epistemic flexibility—which involves statistical learning—is distinct from executive flexibility. In fact, there is considerable research that unlike executive function, which is taxing, effortful and activates regions in the prefrontal cortex (Bunge & Zelazo, 2006), statistical learning is largely automatic and recruits different neural regions depending on the task (Schapiro & Turk-Browne, 2015).

Moreover, there is data which suggests that in some circumstances younger children may actually display more epistemic flexibility than older children and adults, in spite of their reduced executive function. In several tasks, as we noted above, younger children were better than older children and adults at making an initially unlikely inference about abstract causal structure (Lucas et al. 2014, Gopnik et al., 2017) and at creatively solving insight problems (German & Defeyter, 2000). Interestingly this was also true for children of low-income families. Although poverty and poverty-related stressors are generally associated with lower executive functioning abilities for young children (Lawson, Hook, & Farah, 2018), children from low-income families detected statistical regularities in the environment and revised their beliefs to appropriately match the data in much the same way that children in higher-income groups did in spite of lower performance on executive function tasks (Wente et al., 2019).

In this paper, we directly compare 3-year-olds' executive function with their belief revision. If epistemic flexibility is independent from top-down, goal-directed executive flexibility, then we might expect a dissociation in their performance. Conversely, executive function may actually facilitate belief revision. If so, then three-year-olds with low levels of executive function should also fail to revise their beliefs despite the evidence, and there should be a correlation between the two types of performance. Or, as noted executive function might actually be negatively correlated with belief revision. In Experiments 1 and 2, we compared children's performance on a belief revision task to their performance on standard executive function tasks. In Experiment 3, we removed the physical causal framework of the belief revision task to more closely parallel the executive function tasks, again asking whether children change in response to new evidence.

3.2 Experiment 1

The sample size, experimental procedure, and statistical analyses for Experiment 1 and 2 were pre-registered together as separate conditions (http://aspredicted.org/blind.php?x=iq3xe3). Data, however, were collected sequentially and thus are reported as separate experiments. Deviations from the reported plan are reported.

3.2.1 Method

3.2.1.1 Participants

Forty-eight 3-year-olds (M_{age} = 3.58; range = 2.96-3.99; 26 females), who were recruited from and tested at local preschools and museums in the San Francisco Bay Area, participated in this study. An additional 15 children participated but were excluded from analyses for reasons described in our pre-registration: refusal to complete the belief revision task (6), equipment malfunction (4), experimenter error (4), or failure on the pre-test memory aid questions (1): One child was also excluded because of parental inference. We did not include this exclusion criterion in our pre-registration but did so for Experiment 3.

3.2.1.2 Materials

The belief revision task included ten toys (6” x 6” x 1.5”) and 30 blocks (1.5” x 1.5” x 1.5”). Each toy was covered in brown paper and glued to its top was a flat wooden plaque (4.5” x 4.5”) that differed from one another in both color and shape. For example, one toy had a blue circle plaque, another toy had a green triangle plaque, and a third had a pink star plaque. Hidden inside of each toy was a wireless doorbell that, when activated by the experimenter, played a short melody that was unique for each toy. For each of the ten toys, there were three corresponding blocks: (1) a *color match*, which matched the plaque in color but not in shape (e.g., a blue square block matched the blue circle toy), (2) a *shape match*, which matched the plaque in shape but not in color (e.g., a purple circle block matched the blue circle toy), and (3) a *distractor*, which matched the plaque in neither shape nor color (e.g., a yellow L-shape block did not match the blue circle toy). There were also ten small photographs—one of each toy—and two simple drawings, one of a happy face and the other of a sad face.

The Dimensional Change Card Sort (DCCS) task consisted of sixteen cards (3” x 3”) that varied in shape (i.e., rabbit or boat) and color (i.e., red or blue). Two of the sixteen cards were target cards and the remaining fourteen were test cards. One target card depicted a red boat that was taped to a small white sorting tray, whereas the other depicted a blue rabbit taped to a second identical tray. The test cards were identical in style to the target cards, but half of the cards were of a red rabbit and the other half were of a blue boat.

The Day-Night task used twenty cards (3” x 3”), ten of which were of a bright sun on a white background and ten were of a moon and three stars set against a black background.

3.2.1.3 Procedure

Children sat across from the experimenter at a table in a quiet corner in their preschool or museum. Each session lasted approximately 15 minutes and included three tasks: (1) a belief revision task, (2) the DCCS task, and (3) the Day-Night task. The order of the tasks was counterbalanced across participants. The color and shape rule were also counterbalanced such that half of the children observed first the color rule followed by the shape rule, whereas the other half observed the opposite pattern. The learning and revision rule in the belief revision task always corresponded with the pre- and post-switch rule in the DCCS task such that, for example, children who, on the belief revision task, observed the color rule followed by the shape rule also sorted first by color and then by shape on the DCCS task.

3.2.1.3.1 Belief Revision Task

The belief revision task was designed to assess children’s higher-order beliefs and was adapted from Kimura and Gopnik (2019) with procedures that were slightly simplified for younger participants. This task was divided into two phases, a learning phase and a revision phase. In the learning phase, children observed a single toy activate by either a matching shape or color block. Afterwards, in the revision phase, they observed four new toys that activated with the opposite pattern. For example, children who, in the learning phase, observed a toy activated by its matching color block, then saw four new toys in the revision phase that were activated by their matching shape blocks and not their color blocks. At the end of each phase, children completed a single test trial. Across the two phases, children completed five training trials and two test trials, resulting in a total of seven trials. Children observed a new toy for each of the seven trials.

Learning Phase. The experimenter began the task by telling children, “Look, here are my toys. Some blocks make my toys play music, and some blocks do not make my toys play music.” The experimenter then placed two small cards—a happy face and a sad face— in front of the child, randomizing the left-right placement between participants. She explained that when a block makes a toy play music, she would place the block next to the happy face. Blocks that do not activate the toys, however, would be placed next to the sad face. The experimenter then asked the child to point to where she should place blocks that make her toys play music, and where she should place blocks that do not. If the child failed to point correctly, the experimenter repeated the instructions and then asked the questions again. One child failed a second time and was excluded from the analyses as described in our pre-registration.

For each trial, the experimenter laid out pictures of her toys and invited the child to select a toy by pointing to its picture. After the child made a selection, the experimenter removed the pictures and placed the selected toy on the table next to either its color or shape block. The experimenter then placed the block on top of the toy, which seemingly activated it, and emphasized the effect by saying, “It turned on! It’s playing music!” She then asked the child to point to where she should place the block—either on the happy or sad face—and gave corrective feedback if needed. After she placed the block next to the correct card, the experimenter reminded the child of the outcome by saying, “Let’s remember that this block made this toy play music,” as she pointed back and forth between the block and the toy. The experimenter then removed the toy, the block, and the two cards and immediately administered the first test trial.

The test trial began with the experimenter once again asking the child to select a new toy from her set of pictures. The experimenter then placed the selected toy on the table and randomly placed three blocks in a row—the color match, the shape match, and the distractor—on a tray in front of the toy. She then asked the child, “Can you point to what you think will make this toy play music?”, avoiding singular and plural markers, and slid the tray towards the child. After the child responded, the experimenter provided a neutral response, removed the toy and its blocks, and proceeded to the revision phase.

Revision Phase. The revision phase was identical to the learning phase with the exception that for each of the four training trials the experimenter presented both the shape block and the color block, demonstrating that blocks corresponding with the learning rule failed to activate the toys whereas blocks that supported the opposite rule succeeded. For example, children who observed the color rule in the learning phase saw four trials in the revision phase that supported the shape rule and not the color rule. After the four revision trials, the experimenter administered a final test trial to measure whether the child revised or maintained their higher-order belief.

Coding. Responses were coded as being either consistent or inconsistent with the learning and revision rules. For example, children who observed the color rule and selected the color match in the learning phase were coded as having learned the rule. If, however, they had selected the shape match or the distractor block, then they were coded as having not learned the rule. Children were similarly coded as revising the rule if, during the revision phase, they selected the block that corresponded with the revision rule. The experimenter recorded responses during the task. A second coder, blind to the experimental hypothesis, coded a random subset from video recordings (65%). Interrater reliability between the two coders was 98.3%. A third coder, also blind to the hypothesis, resolved any disagreements.

3.2.1.3.2 Dimensional Change Card Sort (DCCS) Task

The DCCS is a widely used measure of executive function to assess children's ability to switch from one task to another (Zelazo, 2006; Zelazo et al., 2003). The experimenter began by placing two sorting trays in front of the child with the left-right placement of the target cards randomized between participants. She then invited the child to sort her cards by either shape or color and demonstrated with two test cards. The child was asked to sort six test cards along the same dimension. Afterwards, the child was instructed to sort six more test cards along the other dimension. The order of the pre- and post-switch rules for the DCCS matched the order of the learning and revision rules for the belief revision task. For example, children who observed the color rule in the learning phase and the shape rule in the revision phase were instructed in the DCCS to sort the cards first by color and then by shape.

Performance on the DCCS was measured by the number of cards that were correctly sorted after the rule switched. Children who correctly sorted five or more post-switch trials were coded as passing the DCCS, whereas those who sorted at least five post-switch trials incorrectly were coded as failing. Those who failed at least five pre-switch trials were excluded from analyses with the DCCS. Once again, the experimenter coded children's responses during the task, and a second coder scored a random subset of participants (65%). Interrater reliability on the DCCS was 98.3%, with a third coder resolving disagreements.

3.2.1.3.3 Day-Night Task

The Day-Night task, another widely used measure of executive function, is a simplified version of the Stroop test that is designed to measure children's inhibitory control (Gerstadt et al., 1994). In this task, the child was first shown a card with a sun and was instructed to say "night." The experimenter then presented a card with a moon and told the child to say "day." Afterwards, the child completed two practice trials, one for each type. If the child responded correctly on both trials, the experimenter immediately administered 16 test trials, 8 of each type, in a pseudo-randomized order. If, however, the child responded incorrectly on either practice trial, the experimenter repeated the instructions and continued administering the practice trials until the child responded correctly to both.

Performance was coded as the number of correct responses on all 16 trials. A primary coder scored children's responses, while a second coder scored a random subset of participants (65%). Interrater reliability between the two coders was 99.7%. A third coder resolved discrepancies.

3.2.2 Results

Preliminary analyses indicated that there were no reliable effects of task order or of rule type on children's performance, *p*s > .05. In the belief revision task, children were just as likely to learn the color rule and revise to the shape rule as they were to learn the shape rule and revise to the color rule. Similarly, in the DCCS they were just as likely to switch from sorting by color to shape as the reverse. As a result, analyses were collapsed across the two rules and the task orders. All analyses were two-tailed.

We first examined performance on the belief revision task by comparing the proportion of children who, on the first test trial, selected the block that corresponded with the learning rule.

If children correctly inferred the higher-order rule, then they should reliably select the block that matches the learning rule. The majority of children indeed chose the block that matched the learning rule (30 of 48 children, 62.5%), selecting this block significantly more often than the block that matched the revision rule (11 of 48 children; 22.9%) and significantly more often than the distractor block (7 of 48 children; 14.6%), $\chi^2(2) = 17.43$, $p < .001$. Comparisons to chance performance (i.e., .33) further revealed that the proportion of those who selected the learning rule block was well above chance, $p < .001$, binomial test, whereas the proportion of those who selected the distractor block was well below chance, $p = .003$ binomial test. This suggests that even after observing just a single training trial, children reliably selected the learning rule block and actively avoided blocks that matched on neither shape nor color. By contrast, children who selected the revision rule block did not differ from chance, $p = .081$, binomial test. Although the data suggest that 3-year-olds, as a group, were able to infer the appropriate higher-order rule, this task was nevertheless challenging as 18 out of 48 children (37.5%) failed to infer the correct rule.

To test whether children then revised their higher-order beliefs after observing four trials of counterevidence, we restricted our analyses to the 30 participants who had demonstratively learned the rule and examined their performance on the second test trial. If, given the new evidence, children revised their higher-order beliefs, then they should now select the block that corresponds with the revision rule. If instead they maintained their beliefs, then they should continue to select the block that matches the learning rule. As shown in Figure 3.1, we found that children chose the block that corresponded with the revision rule (18 out of 30 children, 60.0%) significantly more often than the block that matched the learning rule (6 out of 30 children, 20.0%) and significantly more often than the distractor block (6 out of 30 children, 20.0%), $\chi^2(2) = 9.60$, $p = .008$. This preference for the block that corresponded with the revision rule exceeded chance responding, $p = .002$, binomial test, indicating that at least a subset of 3-year-olds reliably revised their higher-order beliefs. Neither the group that selected the block matching the learning rule nor the group that selected the distractor block differed from chance, ps = .084.

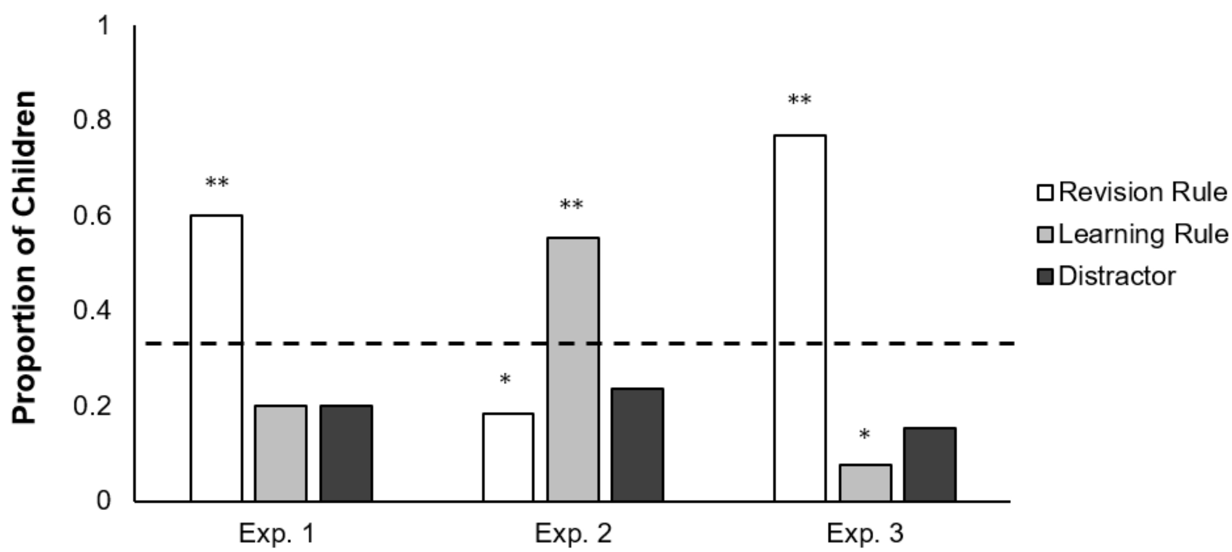

Figure 3.1 Proportion of children who selected the block that matched the revision rule, the learning rule, or neither in the revision or post-switch phase for Experiments 1, 2, and 3.

We then asked whether children's performance on the belief revision task was related to their executive function performance. Nine of the 48 children (18.8%) failed the practice trials for Day-Night even after repeated training. Failure to pass the training for Day-Night was unrelated to their performance on the belief revision task in the learning phase, $p = .064$, McNemar test, and in the revision phase, $p = .143$, McNemar test. Two children (4.2%)—one who had learned the higher-order rule on the belief revision task and one who had not—failed at least five of the pre-switch trials on the DCCS. Data from these participants were, therefore, excluded from further analyses for each respective executive function task. Performance on the DCCS and Day-Night as a function of performance on the belief revision task is depicted in Table 3.1 and Figure 3.2, respectively.

Block Selected	Experiment 1		Experiment 2	
	Passed	Failed	Passed	Failed
Revision Rule	7	8	4	2
Learning Rule	2	4	8	11
Distractor	4	1	4	5
Total	13	13	16	18

Table 3.1 Number of children who passed or failed the DCCS organized by the block selected in the revision phase of Experiments 1 and 2.

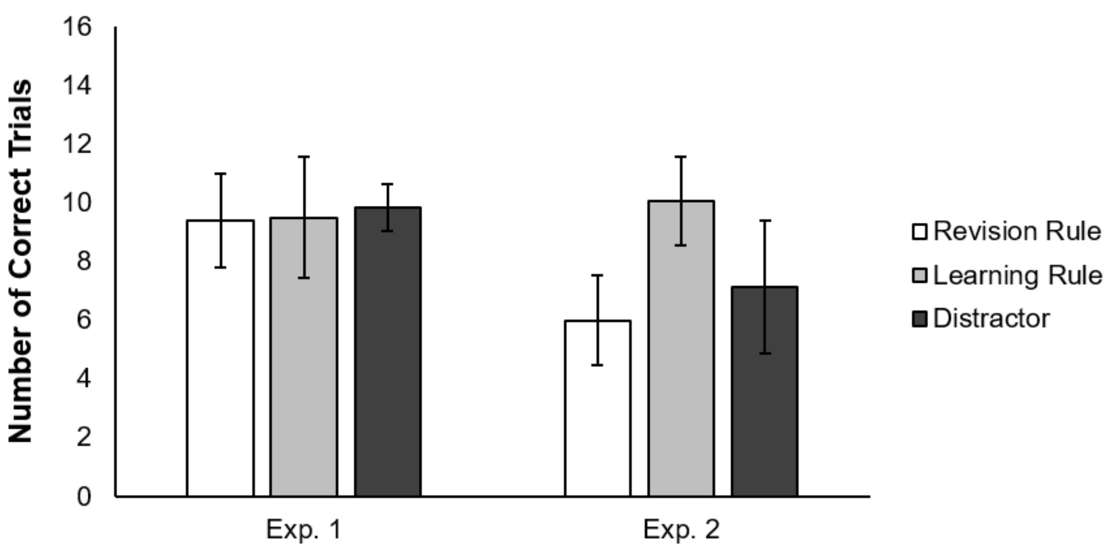

Figure 3.2 Mean accuracy on Day-Night for children who selected the block that matched the revision rule, the learning rule, or neither in the revision phase of Experiments 1 and 2. Error bars represent standard errors of the mean.

Three-year-olds performance on the DCCS and Day-Night tasks in our experiment was similar to their performance in previous studies. Overall, children did not perform better on the DCCS than chance (*M* = 3.21, *SD* = 2.90), $t(46) = .504$, $p = .617$, but performance was bimodally split: 24 of 43 children (55.8%) passed the DCCS (correctly sorted 5 or more post-switch trials), whereas 19 of 43 children (44.2%) failed (incorrectly sorted at least 5 post-switch trials). Only three participants (7.0%) sorted the post-switch trials without a clear pattern and could not be coded. Notably, however, performance on the DCCS was independent of performance on higher-order learning and revision. Children who passed the DCCS responded similarly to those who failed the DCCS on both the learning test trial, $\chi^2(2) = 1.65$, $p = .438$, and revision test trial, $\chi^2(2) = 2.53$, $p = .282$. There was not even a suggestion of either a negative or positive correlation between the two types of tasks.

The Day-Night task was similarly challenging for 3-year-olds. Of those who passed the Day-Night training, children on average responded correctly to 9.05 out of 16 trials (56.6%; *SD* = 4.55). Performance was stable over the course of the task, with children responding with as much accuracy on the first four trials (*M* = 2.59, 64.8%; *SD* = 1.41) as on the last four trials, (*M* = 2.33, 58.3%; *SD* = 1.68), $t(38) = .874$, $p = .387$.

To statistically test whether children's executive function predicted their performance on the belief revision task, we performed two binary logistic regressions using the DCCS bimodal classification, accuracy on Day-Night, and age as predictors of higher-order learning and revision. Both models—the first one predicting whether children learned the higher-order rule and the second one predicting whether they appropriately revised—were not significant, *p*s = .365 and .923, respectively. These results suggest that forming and revising higher-order inferences may be independent from emerging top-down, executive functions.

3.2.3 Discussion

Experiment 1 demonstrates that 3-year-olds learned and flexibly revised their higher-order beliefs in ways that were consistent with the evidence. When participants saw a toy follow the color rule, for example, a majority inferred that a different toy would similarly follow that rule. Afterwards, when they observed several new toys following the shape rule, children revised their higher-order belief. Now, when they were asked to activate yet another toy, a majority of these children selected the shape block.

Moreover, these children showed the typical level of inflexibility in their executive function performance. It is difficult to directly compare the level of performance of children on the two types of tasks given the differences in both the number of trials and exclusion criteria, but in both cases there was considerable variability in performance among the 3-year-olds. However, this variability was not correlated across the tasks, and overall, children's emerging executive functions were unrelated to both the formation and the revision of their higher-order beliefs.

Although it is clear that children revised their beliefs in response to the data, it is unclear whether these revisions reflected a normative interaction between the strength of the initial evidence and the weight of the counterevidence. Indeed, it is possible that children succeeded on this task simply because they were repeating the last observation. That is, children might select the block that corresponds with the revision rule not because the counterevidence was sufficiently strong to outweigh the prior belief but instead because of a recency effect.

Experiment 2 addresses this alternative explanation by repeating Experiment 1 with four learning trials and one revision trial. If children normatively revise their higher-order beliefs, then they should maintain their initial belief despite the revision phase and continue to select the block that is consistent with the learning rule. If, however, they are imitating the last observation, then they should select the block that supports the revision rule. More generally, this experiment acts as a control for other explanations for the performance in Experiment 1 since the only difference between the two tasks is the strength of the evidence. As in Experiment 1, we also examined the relation between higher-order belief formation and revision and executive function.

3.3 Experiment 2

3.3.1 Method

3.3.1.1 Participants

Fifty-nine 3-year-olds (M_{age} = 3.49; range = 2.93-3.99; 33 females) participated. Children were recruited from the same population as was used in Experiment 1. Following the criteria outlined in our pre-registration, six additional children participated but were excluded due to refusal to complete the belief revision task (3), equipment malfunction (2), or experimenter error (1). One child was also excluded due to parental interference. Although we did not pre-register exclusions for parental or sibling interference, we did so for Experiment 3.

Nine children failed the pre-test memory aid questions, either by responding incorrectly on at least one question (7) or by refusing to respond (2). Although we initially created these questions to measure children's understanding of the memory aids and had even pre-registered our analyses to exclude those who failed, preliminary analyses comparing the performance of those who passed the pre-test questions against those who failed yielded no significant differences. Children were equally likely to learn and revise their higher-order rules irrespective of whether, during the introduction of the task, they responded correctly when asked where to put blocks that do and do not make toys play music, *p*s > .05. Given that the experimenter provided corrective feedback throughout the task and reminded the child of the outcome at the end of each trial, it is perhaps unsurprising that we found no effect of the pre-test memory aid. We, therefore, included these participants in the analyses to increase the power and, for Experiment 3, removed this exclusion criteria from our pre-registration.

3.3.1.2 Materials and procedure

The materials and procedures in Experiment 2 were identical to those used in Experiment 1 with the exception that in the belief revision task children observed four learning trials and one revision trial.

As in Experiment 1, the primary experimenter coded children's responses during the task. A random subset of participants were also scored by a second coder (69%). Interrater reliability between the two coders was 98.6% on the belief revision task, 98.6% on the DCCS task, and 97.3% on the Day-Night task. A third coder, also blind to the hypothesis, resolved any disagreements between the first two coders.

3.3.2 Results

Preliminary analyses revealed no reliable effects of rule type on children's performance in the belief revision task and the DCCS, suggesting that children were just as likely to learn and sort along the color rule as they were along the shape rule. The rule type was, therefore, collapsed across participants and is excluded from further analyses. We also found no main effect of task order on children's performance on the belief revision task or on the Day-Night task, *p*s > .05. We did, however, detect a difference in their performance on the DCCS, χ^2 (2, N = 55) = 6.403, *p* = .041. Post-hoc comparisons revealed that children who completed the DCCS 2nd performed significantly worse (11 of 14 children failed the DCCS post-switch; 78.6%) than those who did it 3rd (7 of 20 children failed the DCCS post-switch; 35.0%), *p* = .017, Fisher's exact test, and that this difference in performance was independent of the preceding task. All analyses were two-tailed.

As in Experiment 1, we first considered whether children learned the appropriate higher-order rule by analyzing their block selections on the first test trial. Children in Experiment 2, who had observed four learning trials rather than one, selected the block that matched the learning rule (38 of 59 children; 64.4%) more often that predicted by chance (i.e., .33), *p* < .001, binomial test, and more often than either the block that matched the revision rule (13 of 59 children, 22.0%) or the distractor block (8 of 59 children, 13.6%), $\chi^2(2) = 26.27$, *p* < .001. Children selected these other two blocks significantly below chance (i.e., .33), *p* = .041 and *p* < .001, binomial test, respectively. One child selected all three blocks, a response that reflects the alternative belief that every block, regardless of its shape or color, activates the toys. These findings, which replicate Experiment 1, highlight the fact that 3-year-olds inferred the appropriate higher-order rule from the data. Notably, in addition, children in Experiment 1 who observed one learning trial performed just as well as those who, in Experiment 2, observed four learning trials, *p* = .842, Fisher's exact test.

We next examined whether children revised their higher-order beliefs in response to a single trial of conflicting evidence. Similar to Experiment 1, we restricted our analyses to those who, in the learning phase, demonstrated that they had learned the higher-order rule. We then calculated the proportion of those children who, in the revision phase, selected the block that corresponded with the learning rule, the revision rule, or neither. If children draw normative inferences, then they should continue to select the block that corresponds with the learning rule since they observed more evidence in the learning phase than the revision phase. If, however, children are simply responding because of recency effects, then they should select the block that matches the revision rule.

As expected, when 3-year-olds observed four learning trials followed by one revision trial, they chose the block that matched the learning rule (21 out of 38 children; 55.3%) significantly more often than the block that matched the revision rule (7 out of 38 children; 18.4%) and significantly more often than the distractor block (9 out of 38 children; 23.7%), $\chi^2(2) = 9.30$, *p* = .010 (see Figure 3.1). These responses differed from chance, as children selected the block that matched the learning rule reliably above chance, *p* < .01, binomial test, and, by contrast, selected the block that matched the revision rule reliably below chance, *p* = .033, binomial test. Children who selected the distractor block performed no differently from chance, *p* = .137, binomial test.

Comparisons between Experiment 1 and 2 revealed that children chose the block that matched the learning rule more often in Experiment 2 than in Experiment 1, *p* = .006, Fisher's exact test. The opposite pattern, however, emerged with respect to the block that matched the revision rule: children in Experiment 1 selected the revision rule block more often than those in Experiment 2, $p = .001$, Fisher's exact test. Taken together, this pattern of responding suggests that 3-year-olds, like 4-year-olds, are indeed rational when considering new, conflicting evidence. Children flexibly revise their beliefs but only do so when the counterevidence sufficiently outweighs the prior evidence.

Next, we explored whether children's performance on the learning and revision task was related to their executive functions. Children in Experiment 2 performed like those in Experiment 1 on Day-Night and the DCCS. Fifteen of the 59 children (25.4%) failed the Day-Night practice trials despite repeated training, and one parent interfered while the experimenter administered the task. Performance on the Day-Night training trials was unrelated to whether children learned, $p = .383$, McNemar test, or maintained their higher-order rules, $p = .064$, McNemar test. Two children failed the DCCS pre-switch trials (3.4%), one who had learned the higher-order rule and one who had not. Their data were excluded for each respective task.

Children's performance on the DCCS was once again no different from chance (M = 2.96, SD = .290), $t(56) = -.091$, $p = .928$ (not pre-registered), and responses were bimodally distributed in a similar way, with nearly half of the children either correctly (27 of 57 children, 47.4%) or incorrectly (28 of 57 children; 49.1%) sorting 5 or more trials in the post-switch phase. Two children had an unclear sorting pattern and were unable to be classified (3.5%). Similar to Experiment 1, children who passed the DCCS performed similarly on the belief revision task to those who failed the DCCS on both the learning test trial, $\chi^2(2) = 1.60$, $p = .449$, and revision test trial, $\chi^2(2) = 1.14$, $p = .566$ (see Table 1). There was no significant difference on the DCCS between experiments, $p = .687$, Fisher's exact test.

On the Day-Night task, children responded correctly on 8.16 out of 16 trials (51.0%; SD = 5.67). Similar to Gersdhat et al. (1994), performance declined over the course of the task with children responding with greater accuracy on the first four trials (M = 2.72, 68.0%; SD = 1.52) compared to the last four trials (M = 1.79, 44.8%; SD = 1.61, $t(42) = 4.06$, $p < .001$. Children's performance in Experiment 2 did not differ from that in Experiment 1, $t(80) = .778$, $p = .439$.

We then performed two binary logistic regressions to test the relation between executive function and belief revision. Using age as well as performance on the DCCS and Day-Night as predictors of learning and revising higher-order beliefs, we found the two models were not significant, ps = .590 and .378, respectively.

Age Comparison to Kimura and Gopnik (2019). Finally, we compared the 3-year-olds on the belief revision task in Experiment 1 and 2 to the 4 and 5-year olds in Kimura and Gopnik (2019), who completed a similar evidence-based belief revision task (analyses were not pre-registered). Even though older children in general have greater executive function skills the 3-year-olds were just as likely to revise their beliefs (18 out of 30 children, 60.0%) as 4- and 5-year-olds (19 out of 30 children; 63.3%), $p = 1.00$, Fisher's exact test, when shown one learning trial and four revision trials. When the evidence was reversed and they were shown four learning trials and one revision trial, both the younger (21 out of 38 children; 55.3%) and older children (22 out of 32 children; 68.8%) maintained their beliefs, $p = .326$, Fisher's exact test.

3.3.3 Discussion

The findings from Experiment 2 support the interpretation that 3-year-olds revise their higher-order beliefs by considering the normative interaction between the initial evidence and the counterevidence. When children observe four learning trials followed by one revision trial, they continue to select the block that corresponds with the learning rule despite having just observed counterevidence. This suggests that their performance in Experiment 1 was not driven by a recency effect or other superficial effects—as children in Experiment 2 failed to repeat the recently observed revision rule—but was instead driven by the strength of the evidence, particularly the strength of the counterevidence to outweigh the prior evidence.

Consistent with our findings in Experiment 1, young children's executive function fails to predict whether they revise or maintain their higher-order beliefs. Although both tasks—the belief revision task and the DCCS—require a switch from a color rule to a shape rule, there are a number of other features of the tasks that differ. In particular, the belief revision task involves a judgment about the physical causal structure of the external toy, based on evidence, whereas the DCCS involves following an arbitrary social rule generated by the experimenter. In other tasks, children perform substantially better in a causal context than a non-causal one (e.g., Goddu, Lombrozo, & Gopnik, 2020, Walker & Gopnik, 2014).

In the following experiment, we continue to test the hypothesis that 3-year-olds are epistemically flexible by creating a non-causal, higher-order rule-switch task that provides the child with evidence, but more closely matches the DCCS task in other ways. If children are epistemically flexible because they observe evidence, then they should select the block that matches the post-switch rule even when the evidence involves a social rule rather than a physical causal principle. If, however, children are flexible because of the physical causal framework, then they might perseverate, similar to the DCCS, and continue to select the block that matches the pre-switch rule.

3.4 Experiment 3

3.4.1 Method

The sample size, experimental procedure, and statistical analyses were pre-registered (see https://aspredicted.org/blind2.php). Deviations from these plans are reported.

3.4.1.1 Participants

Sixteen 3-year-olds (M_{age} = 3.31; range = 2.93-3.83; 8 females) were recruited from the same population as in Experiments 1 and 2. No participants were excluded. Due to the unexpected emergence of the pandemic, data collection abruptly stopped, leaving our sample size smaller than pre-registered.

3.4.1.2 Materials and procedure

The materials and procedure in Experiment 3 were identical to those used in the belief revision task of Experiment 1 with the exception that the experimenter removed the physical causal framework and instead invited children to play a matching game. The task was divided into two phases: a pre-switch phase and a post-switch phase, which corresponded with the learning and revision phases in Experiment 1, respectively. In the pre-switch phase, children saw a single toy that matched with either a color block or a shape block. In the post-switch phase, they observed four new toys that matched along the opposite pattern. Similar to Experiment 1, both phases ended with a single test trial to demonstrate whether children applied the pre- or post-switch rule to a new toy. Across the two phases, children completed five training trials and two test trials.

Pre-Switch Phase. Similar to Experiment 1, the experimenter began the task by introducing the child to her matching game, explaining that some blocks match her toys whereas other blocks do not. She then presented two small cards, one of a happy face and the other of a sad face, and said that matching blocks are to be placed next to the happy face, whereas non-matching blocks are to be placed next to the sad face. The child then demonstrated their understanding of the placement cards by pointing to where the experimenter should place matching and non-matching blocks. If the child pointed incorrectly, the experimenter provided corrective feedback and repeated the instructions before advancing to the first training trial.

The training trial began with the experimenter presenting pictures of her toys and instructing the child to select a toy by pointing to its picture. After the child made a selection, the experimenter then placed the selected toy and its corresponding block next to each other on the table. Similar to Experiment 1, she then placed the block on top of the toy but, unlike Experiment 1, the toy did not play music. Instead, the experimenter exclaimed, “It matches! This block matches my toy.” She then asked the child where she should place the block, giving corrective feedback if needed, and placed it next to the correct card. After reminding the child that the block matched the toy, the experimenter removed the toy and its block as well as the two cards and proceeded to the test trial.

The test trial in Experiment 3 was identical to the test trial in Experiment 1 with the exception that children were asked to select which object matched the new toy. Once the child responded, the experimenter provided a neutral response and advanced to the post-switch phase.

Post-Switch Phase. The post-switch phase was similar to the pre-switch phase, but the experimenter provided positive and negative evidence that supported the opposite rule. For each trial, the experimenter showed two blocks—the color and the shape block—and, after placing each one on top of the new toy, said that the block that corresponded with the pre-switch rule did not match the toy whereas the other block did. The order in which the two blocks were presented was randomized across trials. After placing each block next to the appropriate card, the experimenter reminded the child of the outcome, removed the toy and its blocks, and repeated the procedure for the remaining three post-switch trials. After the fourth training trial, the experimenter administered the final test trial. Similar to Experiment 1, the experimenter neither stated the pre- and post-switch rules nor highlighted the switch.

Coding. Block selections for each test trial were coded as being either consistent or inconsistent with the pre- and post-switch rules. The experimenter recorded responses during the experiment, whereas a second coder coded a random subset from video recordings (81.2%). Interrater reliability was 100%.

3.4.2 Results

Preliminary analyses yielded no reliable effects of rule type, suggesting that children were just as likely to match by the color rule as they were the shape rule. The sorting rule was, therefore, collapsed. All analyses were two-tailed.

On the first test trial, children overwhelmingly selected the block that corresponded with the pre-switch rule (13 of 16 children; 81.3%), doing so at a rate that exceeded chance performance (i.e., .33), $p < .001$, binomial test. More children selected this matching pre-switch block than either the matching post-switch block (3 of 16 children, 18.8%) or the distractor block (0 of 16 children), $\chi^2(1) = 6.25$, $p = .012$. As in Experiment 1, children quickly learned the matching rule from a single non-causal event even without explicit instructions.

Of interest is how these thirteen children responded on the second test trial, after they observed four post-switch training trials. When asked to match a new toy with its blocks, the majority of children switched and selected the block that matched the post-switch rule (10 out of 13 children; 76.9%). Children selected this post-switch block over the matching pre-switch block (1 out of 13 children, 7.7%) and over the distractor block (2 out of 13 children, 15.4%), $\chi^2(2) = 11.23$, $p = .004$. A similar pattern emerged when we compared their performance to chance. Children chose the block that matched the post-switch block significantly above chance, $p = .001$, binomial test, and chose the block that matched the pre-switch block significantly below chance, $p = .039$, binomial test. The proportion of those who selected the distractor block did not differ from chance, $p = .139$, binomial test. Together this suggests that children reliably switched from one rule to the other after observing four non-causal post-switch trials.

As a separate analysis, we compared performance on the non-causal task in Experiment 3 with performance on the causal task in Experiment 1 (analyses were not pre-registered). We found no significant difference in their block selections on either the first test trial, $\chi^2(2) = 3.42$, $p = .181$, or the second, $\chi^2(2) = 1.35$, $p = .510$, suggesting that children switch their higher-order rule irrespective of whether they observed evidence for a physical causal relationship or a social rule.

3.4.3 Discussion

The findings from Experiment 3 parallel those from Experiment 1. Children, regardless of whether they observed physical causal evidence (Experiment 1) or rule-based evidence (Experiment 3), learned the appropriate higher-order rule from a single observation and switched to a different higher-order rule when shown enough counterevidence.

Unlike in the DCCS, where the experimenter simply instructs children to sort by a color or shape rule, this task provides children with clear evidence in support of a given higher-order rule. Although the experimenter highlights the relevant dimension in the DCCS—telling children to sort by color, for example—children are only provided with two examples, for example, a blue rabbit and a red boat, whose correct placement switches when they are sorted by the other dimension. The only evidence children observe is that of the experimenter demonstrating the pre-switch rule at the beginning of the task. There is no evidence supporting the post-switch rule, just the experimenter’s instruction.

Moreover, in the DCCS children must sort identical cards in two different ways, creating an arbitrary conflict that may be challenging for young learners to override. We eliminated this conflict by introducing different toys with different blocks that together supported a specific higher-order rule. Children had to generalize the rule to a new toy, rather than treating the same card in two different ways. Children may have succeeded on this task because we eliminated the conflict. Indeed, it may be that perseverations occur in the DCCS precisely because children have difficulty overcoming conflict (Diamond, Carlson, & Beck, 2005; Kloo & Perner, 2005). Further research is necessary to determine which of these features of the task, the provision of evidence or the elimination of conflict was most important.

3.5 General Discussion

Taken together, these three experiments reveal that 3-year-olds, like 4-year-olds, rationally update their beliefs by weighing the statistical strength of the evidence and that this flexibility is independent of children's emerging executive functions. When shown evidence that repeatedly violated a weakly held belief, they changed their minds to reflect the new data, doing so irrespective of their executive functions (Experiment 1). Children, however, were not revising needlessly. When they observed a small amount of evidence that conflicted with a strongly held belief, they maintained their belief, suggesting that their revisions were influenced by the strength of the counterevidence (Experiment 2). Even in the absence of a physical causal framework, children flexibly switched between rules when presented with new evidence, sorting first by one social rule and then by another (Experiment 3).

These findings underscore the rationality of children's belief revisions and, notably, highlight the distinction between epistemic flexibility and executive goal-directed, top-down flexibility. Children are remarkably sensitive to new evidence. They can form high-level inferences from a single event and can track statistical regularities in the environment to guide their beliefs. This sensitivity to statistical information can then explain why children change their minds in some context but not in others. Executive function, in contrast, recruits inhibitory control and working memory to deliberately control behaviors. These top-down processes operate together to facilitate goal-directed change, such as adjusting to sort by a new rule. These two forms of flexibility may appear to be similar, but in fact, they may be distinct from each other and recruit different mechanisms.

Moreover, there did not appear to be age related changes between 3- and 4-year-olds on the belief revision task in contrast to the executive function tasks. Given that performance on the belief revision task was similar despite age-related improvement in executive function, and that there was no correlation between performance on the two types of tasks, it seems possible that other powerful learning mechanisms are recruited when tracking evidence and updating beliefs. These findings are consistent with the idea that there is a dissociation between "explore" abilities such as wide-ranging spontaneous learning from data, which appear to be present even in very young children and "exploit" executive function abilities such as planning, focus and goal-directed action, which appear to develop through childhood (Gopnik, 2020).

However, it is also possible that our findings reflect other features of the executive function tasks, and particularly the DCCS, that differ from the belief revision tasks beyond the executive versus epistemic contrast. Experiment 3 suggests that the physical causal nature of the task, by itself, does not explain the dissociation between belief revision and executive function. However, children may have succeeded on the belief revision task because the task reduced the conflict inherent to bi-dimensional cards. Unlike the standard DCCS, the belief revision task separates the two dimensions so that the matching color block for one toy is not the matching shape block for another. This may eliminate the need to re-describe the objects and reduce the conflict (Kloo & Perner, 2005).

In addition, very young children may have difficulty forming abstract representations, such as color and shape, when the supporting evidence is insufficient. When children sort identical cards first by one rule and then by another, they may form stimulus-specific representations that preclude them from forming more abstract representations (Kharitonova, Chien, Colunga, & Munakata, 2009). For example, children may form the specific rule that red boats belong in one tray while blue rabbits go in another instead of forming the more general rule to sort by color. Later when they are explicitly asked to switch their general rule, children may fail, especially in the absence of any evidence that the new rule is correct. Four-year-olds appear to have little difficulty forming abstract, higher-order representations when they are instructed to do so, but 3-year-olds may require direct evidence to form and revise these representations. Indeed, it is possible that the younger children formed the first representation precisely because the experimenter provided a demonstration at the beginning of the task. The importance of providing diverse examples when forming abstract representation can be explained by hierarchical Bayesian learning, which describes how diverse lower-order exemplars inform more abstract insights (Perfors, Tenenbaum, Griffiths, & Xu, 2011; Xu & Tenenbaum, 2007). Returning back to the belief revision task, children who observed different blocks activating different toys may, as a result, have been more likely to form higher-order rules over stimulus-specific ones. Thus the difference between the two tasks may involve the content of the evidence that children saw, in particular the fact that the evidence was varied, as well as the simple fact that they saw counterevidence at all.

From the adult perspective, both the ability to flexibly revise our beliefs in the face of evidence, and the ability to change our behavior arbitrarily, are important cognitive capacities. In fact, from the adult perspective, it may seem intuitively easier to intentionally switch behaviors when you are clearly instructed to do so, than it is to infer a novel higher-order belief. It is surprising both that 3-year-olds ever succeed at the higher-order belief revision task, and that they ever fail on the DCCS. Our results, suggest that although very young children are developing both types of abilities, they emerge on independent trajectories

Knowing when to ignore conflicting evidence and when to update beliefs is a challenging task. Although it may be sensible to have robust beliefs, especially if there are trivial discrepancies in the environment, it is also reasonable to have flexible beliefs in response to new information. Given that the world is dynamic yet stable, it is important that beliefs are flexible yet also robust. Despite 3-year-olds’ limited executive function, our experiments indicate that young children can flexibly revise their beliefs and indeed, like 4- and 5-year-olds (Kimura & Gopnik, 2019), do so rationally. The ability to track statistical regularities may help to inform learners when the evidence reflects an anomaly and when it represents an underlying shift. The fact that beliefs can change in response to new data offers us all comfort that what was may not always be.

Chapter 4

Adults’ inflexibility to learn from new evidence

4.1 Introduction

Young children form and revise their beliefs rapidly and accurately—sometimes more accurately than adults (Gopnik, Griffiths, & Lucas, 2015). For example, when preschoolers and adults are shown a toy machine that lights up when particular blocks are placed on top, both age groups initially believe that individual blocks--and not the block combinations--activate the toy (Lucas, Bridgers, Griffiths, & Gopnik, 2014). That is, for instance, if blocks A and B are placed on the machine together, children and adults think that an individual block (e.g., block A) causes the machine to activate rather than both blocks (e.g., blocks A and B). However, when shown evidence to the contrary (i.e., that only *combinations* of blocks activate the machines, whereas individual blocks do not), children are more likely than adults to revise their initial belief in favor of this alternative and, critically, to extend this to new blocks. This pattern even holds in non-Western cultures where people engage in more holistic as opposed to analytic reasoning (Wente et al., 2019), and across the childhood as preschoolers outperform school-aged children, who in turn outperform adult (Gopnik et al., 2017).

This developmental decline with older learners performing worse than younger ones is consistent even at younger ages. In particular, toddlers are more likely than preschoolers to correctly revise their beliefs about higher-order relational properties, and preschoolers are more likely than grade-schoolers to correctly revise their beliefs about higher-level social principles. For instance, when children are shown a toy machine that plays music when two blocks of the same kind (or different kinds) are placed on it, 18-month-olds overwhelmingly learn the key relational patterns—that is, *same* blocks activate the machine or, in contrast, that *different* blocks activate the machine. Thirty-month-olds, however, fail to learn these higher-order relations, focusing instead on shallow perceptual features (Walker & Gopnik, 2014). Similar patterns have been observed in children’s understanding of social relations. Four-year-olds correctly infer higher-order social relations from evidence whereas six-year-olds often rely on incorrect, unsupported biases (Seiver, Gopnik, & Goodman, 2013). This developmental decline, although consistent across several tasks, is nonetheless surprising given that performance on cognitive tasks usually improves with age. Why in these cases do younger children reliably outperform older children and adults?

There are at least two possibilities, both of which are well explained by Bayesian learning. Given that children have less knowledge than adults, they necessarily have flatter priors and, therefore, have less reason to strongly believe in any one belief. Thus, one possibility for why children outperform adults is that children, because of their weaker priors, may require less counterevidence to revise their beliefs. A second possibility, which is not necessarily mutually exclusive from the first, is that children may be specifically better at revising to *unusual* hypotheses because of their intrinsic flexibility to consider alternative hypotheses. That is, children may be searching through their hypothesis space more widely than adults and, as a result, consider a broader range of possibilities. With this wider search, children may then be more likely to evaluate and revise to unusual hypotheses. In other words, younger children may be employing a strategy that favors wide search patterns with radical changes, whereas adults may by using a strategy that leads them to narrow searches with minor revisions. Notably, however, this flexibility is helpful when learning unusual hypotheses but potentially less so when learning more typical ones that require minimal change.

Although these two accounts are closely related—i.e., weaker priors should lead to wider search patterns, whereas stronger priors should lead to narrower searches—it is possible that the ability to flexibly consider alternative hypotheses is independent of prior knowledge. This flexibility may be mediated by cognitive control and supported by the prefrontal cortex, with the decline in flexibility reflecting the protracted maturation of the prefrontal cortex. This slow maturation leaves children with notable deficits in their cognitive control and, coincidentally, may promote the uninhibited, flexible reasoning that is advantageous for some types of learning (Thompson-Schill, Ramscar, & Chrysikou, 2009). As the prefrontal cortex matures and the ability to exert top-down control improves, learners may develop strategies and heuristics to promote efficiency at the cost of flexibility. One strategy may be to exploit prior knowledge by generalizing from past experience. These generalizations are particularly useful in stable environments, where the past can inform the present. However in unstable environments these generalizations may have the unintended consequence of prematurely trapping learners (Rich & Gureckis, 2018). Unable to consider alternative perspectives, learners with higher levels of cognitive control may then make misguided inferences compared to learners with lower levels.

Indeed, young children with limited cognitive control show greater flexibility when solving insight problems. Five-year-olds, for instance, can easily repurpose an object for an alternative use to achieve a goal. For example, they overturned a box, using it as a stool, to reach for a toy on a high shelf. Older children and adults, however, struggle to think beyond an object's conventional function, resulting in a functional fixedness that prevents them from reaching a solution (Duncker, 1945; German & Defeyter, 2000). This enhanced flexibility is also evident in children's selective attention. Unlike adults, who were better at selectively attending to goal-related information, preschoolers were better at processing seemingly irrelevant information (Plebanek & Sloutsky, 2017).

This privileged relation between flexibility and the prefrontal cortex is also evident in adults with disrupted or damaged prefrontal cortices. Suppressing cortical activity in the prefrontal cortex with low-level electrical currents similarly improves adults' creative reasoning and enhances their problem solving (Chi & Snyder, 2011, 2012; Weinberger, Green, & Chrysikou, 2017). For example, adults who receive transcranial direct current stimulation (tDCS) in the left prefrontal cortex are faster at generating unusual uses for familiar objects (Chrysikou, Hamilton, Coslett, Datta, Bikson, & Thompson-Schill, 2013) and are better at solving classic insight tasks such as matchstick arithmetic problems and the nine-dot problem (Chi & Snyder, 2011, 2012). Even adult patients with lesions in their lateral prefrontal cortex are surprisingly better than healthy controls at solving problems that require overcoming strong constraints and biases (Reverberi, Toraldo, D'agostini, & Skrap, 2005).

It is possible for healthy adults to reason flexibly without first damaging their prefrontal cortex or undergoing cortical suppression. In particular, adults can simply be primed to reason more creatively by engaging in tasks that require them to think broadly. For instance, adults are less susceptible to functional fixedness if they are first primed to generate unusual uses for common items (Chrysikou, 2006).

Given that adults have strong biases, beliefs, and heuristics that can sometimes lead them astray, we asked whether adults can be encouraged to generate other alternatives that, despite being unusual and uncommon, better explain the data. In particular, we investigate whether we can encourage adults to endorse the conjunctive rule that two or more blocks are needed to activate a novel toy (Lucas et al., 2014). Across three experiments, we demonstrate the challenge in overturning adults' disjunctive bias. In Experiment 1, adults were asked to generate either 1 or 6 descriptions for common objects with the assumption that generating multiple descriptions would require a wider search than generating one. Afterwards, they observed a novel toy activate when particular combinations of blocks were placed on top of it. In Experiment 2, participants observed the unusual conjunctive rule applied across multiple domains and contexts. In Experiment 3, we introduced financial incentives to examine whether adults can be motivated to endorse the conjunctive rule. Together, these experiments provide insight into the challenge of revising adults' robust beliefs.

4.2 Experiment 1

Adults were presented with a list of familiar objects and, for each item, generated either one or six descriptions. Participants then observed a series of causal events that supported the conjunctive principle—i.e., only combinations of blocks activated the machines, whereas individual blocks did not. Afterwards, adults saw an ambiguous activation pattern that was consistent with both the conjunctive and disjunctive principles—i.e., individual blocks activate a machine—and were asked to select those blocks that activated the machine. If adults are primed to think creatively and, therefore, more broadly, then adults should infer the hypothesis that, although unusual, is more consistent with the observed data. Hence, adults who generate six descriptions should be more likely to infer that combinations of blocks activate a toy machine. In contrast, if adults have such strong priors that they cannot be encouraged to consider other hypotheses, then adults should infer the disjunctive rule regardless of whether they generated one or six descriptions.

4.2.1 Method

4.2.1.1 Participants

One hundred and twenty-five undergraduate students from a large public university participated in exchange for course credit (M_{age} = 21.28 years, SD_{age} = 2.31 years, 84 females). Participants were randomly assigned to either the *1 description* (N = 61, M_{age} = 21.26 years, SD_{age} = 1.85 years, 42 females) or the *6 descriptions* condition (N = 64, M_{age} = 21.29 years, SD_{age} = 2.69 years, 42 females). An additional six students participated but were excluded from analyses for experimenter error (4), equipment malfunction (1), or correctly inferring that the experimenter operated the machine with a switch (1).

4.2.1.2 Materials

Two questionnaires, adapted from Christensen and Guilford (1958) and normed by Chrysikou (2006), were created. Both questionnaires asked participants to generate descriptions of common items (e.g., pillow, keys, etc.), but one questionnaire had participants list one description for each item, whereas the other questionnaire had participants list six descriptions. Participants who listed one response for each object considered a total of 72 items, whereas those who listed six responses per object considered only 12 items. Both questionnaires were divided into three sections with either 4 or 24 items per section. Each section lasted no more than five minutes for a maximum total of 15 minutes.

Modeled after the blicket detector paradigm (Gopnik & Sobel, 2001), a toy machine was designed such that it appeared the certain blocks made a toy machine light up and play music whereas other blocks did not. The toy, however, was controlled by a switch surreptitiously operated by the experimenter. Corresponding with the toy were gray nine blocks, each with a different shape, that were organized into three sets of three.

4.2.1.3 Procedure

Adults were tested in groups of up to four and were seated at a table across from the experimenter. Each session lasted approximately 25 minutes and consisted of two tasks: (1) a priming task, and (2) a causal learning task. In the *priming task*, participants were asked to list either one or six descriptions for common, everyday items. The *causal learning task*, in contrast, was adapted from Lucas et al. (2014) and measured adults' abstract causal learning. For each task, participants were given printed packets on which to record their responses.

4.2.1.3.1 Priming task

The priming task was designed to elicit broad reasoning by listing descriptions for objects. For both the 1 description and 6 descriptions conditions, the experimenter began by reading the instructions to participants. In the instructions, participants were told that they would consider common items and that for each item they were asked to list either one or six descriptions depending on the condition. The experimenter then provided a possible response, using a newspaper as an example. For those assigned to the one description condition, the experimenter stated that a description for a newspaper is, "consists of news articles." For those in the six descriptions condition, she repeated the same description but then provided five additional descriptions, each one distinct from the others. In both conditions, participants were encouraged to respond quickly, leaving as few blanks as possible. Participants had 15 minutes to consider 12 items if they were asked to list 6 descriptions or 72 items if they were asked to list only one. The task was divided into three sections, with each section consisting of either 4 or 24 items that lasted 5 minutes.

4.2.1.3.2 Causal learning task

Immediately after the priming task, participants observed the experimenter demonstrating a series of events that supported the conjunctive rule (Lucas et al., 2014). They were then shown an ambiguous pattern of evidence that was consistent with the more typical disjunctive rule and the more unusual conjunctive rule and were asked to infer the causal property by activating the machine.

The experimenter began the task by showing adults a large container with blocks and explaining that the goal of the task is to identify which of the blocks are *blickets*. The experimenter then presented a novel toy and explained that, although blickets cannot be identified by their physical appearance, blickets have *blicketness* inside of them that when placed on top of the toy will make the toy play music. The experimenter then invited participants to figure out which of her blocks were blickets.

The experimenter began the training trial by selecting three blocks from the container, seemingly at random, and placing them in a row directly in front of the toy. She then labeled the objects based on their shape, saying, for example, "Let's call this one Rectangle" for the rectangular block. Next, the experimenter placed the objects, both individually and in combinations, on top of the toy (see Figure 4.1). When placed individually, the blocks failed to activate the toy. However, when placed in a particular combination with another block, the toy activated. To emphasize each outcome, the experimenter would exclaim, "It turned on!" when the toy activated and "It did not turn on!" when it did not. After demonstrating the activation patterns, the experimenter then pointed to the first block and asked, for example, "Do you think Rectangle is a blicket or not a blicket?" She then repeated the question for the remaining two blocks. After participants recorded their responses, the experimenter removed the three blocks and administered the second training trial with three new blocks. The combination of the blocks for each set were counterbalanced across participants.

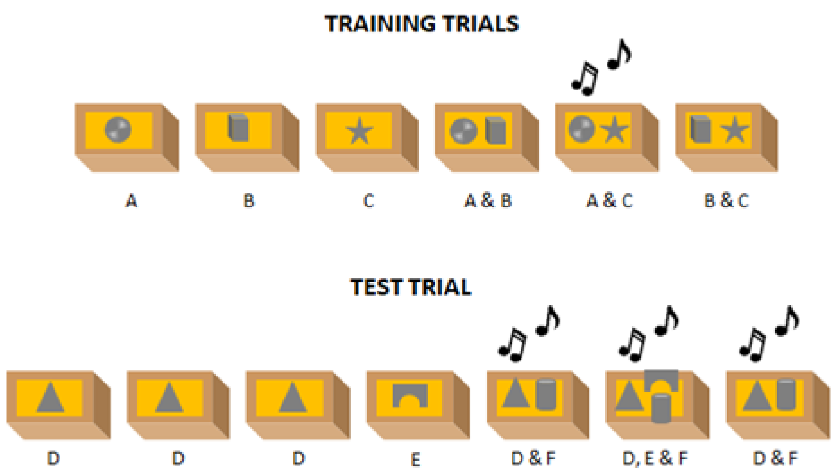

Figure 4.1 Activation pattern in the causal learning task for Experiment 1.

After completing two training trials, the experimenter immediately proceeded to the test trial. Unlike the training trails, however, the experimenter demonstrated an ambiguous activation pattern that could support either the disjunctive or the conjunctive rule. Afterwards, participants were asked whether each object was a blicket. Because participants may be sensitive to subtle linguistic cues that could lead them to interpret "blicket" as an individual cause, we included a second measure that asked to activate the toy by selecting blocks. Singular and plural markings were avoided since these cues could influence whether participants responded conjunctively or disjunctively.

4.2.1.4 Coding

Three independent raters coded the priming responses along two dimensions of creativity: (1) novelty of the response; and (2) appropriateness (see Hennessey & Amabile, 2010; Runco, 2010, for a review of creativity measures). Each dimension ranged on a scale from 0-100, with higher scores indicating greater novelty and appropriateness, respectively. For example, describing a paperclip as an object that "holds papers together" was scored as being less novel than the description that a paperclip "conducts electricity with a battery."

4.2.2 Results

4.2.2.1 Priming task

We found no significant difference in the total number of responses between conditions, suggesting that those who were instructed to list 1 description per item (M = 64.26 out of 72 responses; SD = 10.48) were just as likely to produce a response as those who were instructed to list six (M = 64.26 out of 72 responses; SD = 10.79), $t(123) = -1.15$, $p = .25$. Contrary to our predictions, however, adults who listed one description produced more novel responses (M = 7.80 out of 100; SD = 2.41) than those who listed six (M = 4.75 out of 100; SD = 2.25), $t(123) = -7.34$, $p < .001$ and, similarly, produced more appropriate responses for each item (1 Description: M = 85.54 out of 100; SD = 10.18; 6 Descriptions: M = 56.59 out of 100; SD = 8.45), $t(123) = -17.33$; $p < .001$. Despite these significant differences, responses tended to be neither novel nor unusual as indicated by the low ratings. However, given that adults were not explicitly instructed to produce novel or unusual descriptions, this result was not surprising.

4.2.2.2 Causal learning task

To test whether adults reasoned conjunctively, we examined which objects participants judged to be blickets and then assessed the blocks they selected when asked to activate the toy. If participants reasoned conjunctively, then they should infer that objects D and F are blickets given that the machine always activated when the two objects were placed on the machine together. Likewise, when asked to activate the machine, participants should select multiple objects, particularly D and F. If, however, participants reasoned disjunctively, then they should infer that only object F is a blicket since objects D and E failed to activate the machine individually Similarly, they should select just F when asked to activate the machine. All analyses were two-tailed.

Blicket Judgments. Figure 4.2 shows the mean proportion of blicket judgments by condition for each test block.

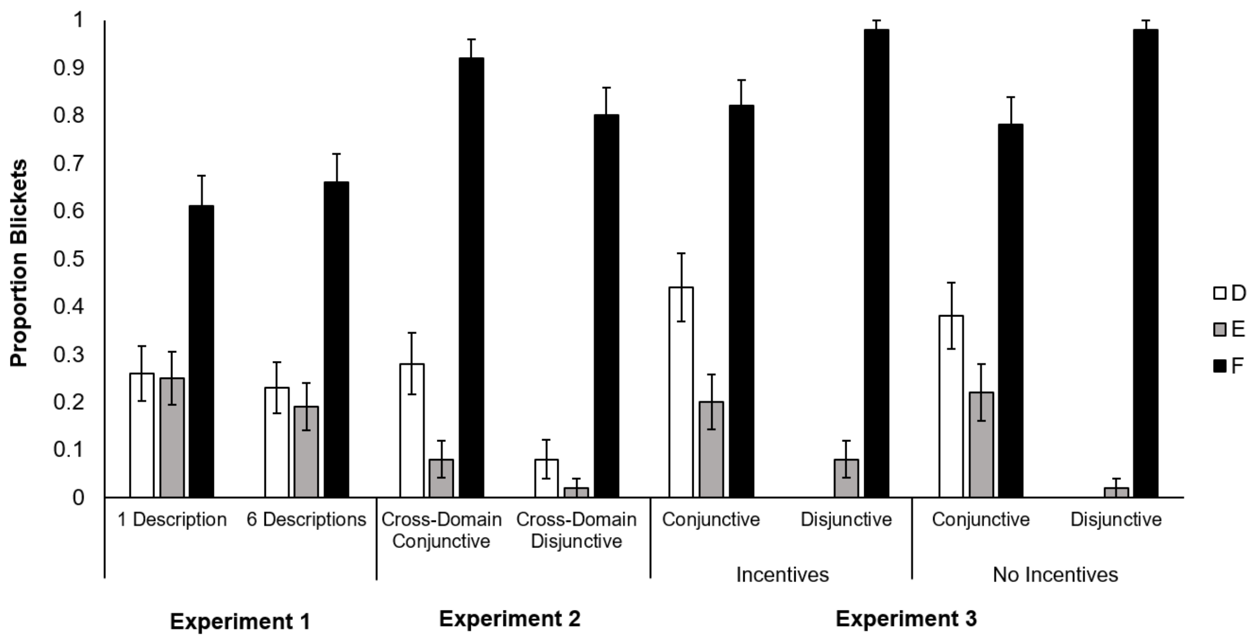

Figure 4.2 Average proportion of participants who judged objects D, E, and F as blickets in Experiments 1, 2 and 3. Error bars indicate one standard error of the mean in each direction.

To test whether adults who listed six descriptions were more likely to infer the conjunctive principle than those who listed one, we performed a series of analyses to examine blicket judgments. Regardless of whether participants listed one or six descriptions, we found that participants in both conditions were more likely to label the unambiguous object F as a blicket (1 Description: M = .61; SD = .49; 6 Descriptions: M = .66; SD = .48) than the conjunctively active object D (1 Description: M = .26; SD = .44; 6 Descriptions: M = .23; SD = .43), p's < .001, McNemar's test, or the uncertain object E (1 Description: M = .25; SD = .43; 6 Descriptions: M = .19; SD = .39), p's < .001, McNemar's test, suggesting that participants believed one object was responsible for activating the machine. There were no significant differences between adults' judgments of objects D and E in either condition.

We then compared the two conditions and found no significant difference in their judgment of D. That is, adults who listed 1 description per item were just as likely to report object D as a blicket as those who listed 6 descriptions, p = .84, Fisher's exact test. Likewise, there were no significant differences in their judgments of E, p = .52, Fisher's exact test, and F, p = .58, Fisher's exact test.

Intervention Choices. As an additional analysis, we examined participants' intervention choice (see Figure 4.3). More specifically, we were interested in: (1) whether participants included object D, which activated the machine in conjunction with another object but never individually; (2) whether they selected multiple objects, which is consistent with the belief that objects activate the machine together; and (3) whether they selected only object F, which is a choice that is consistent with the disjunctive principle.

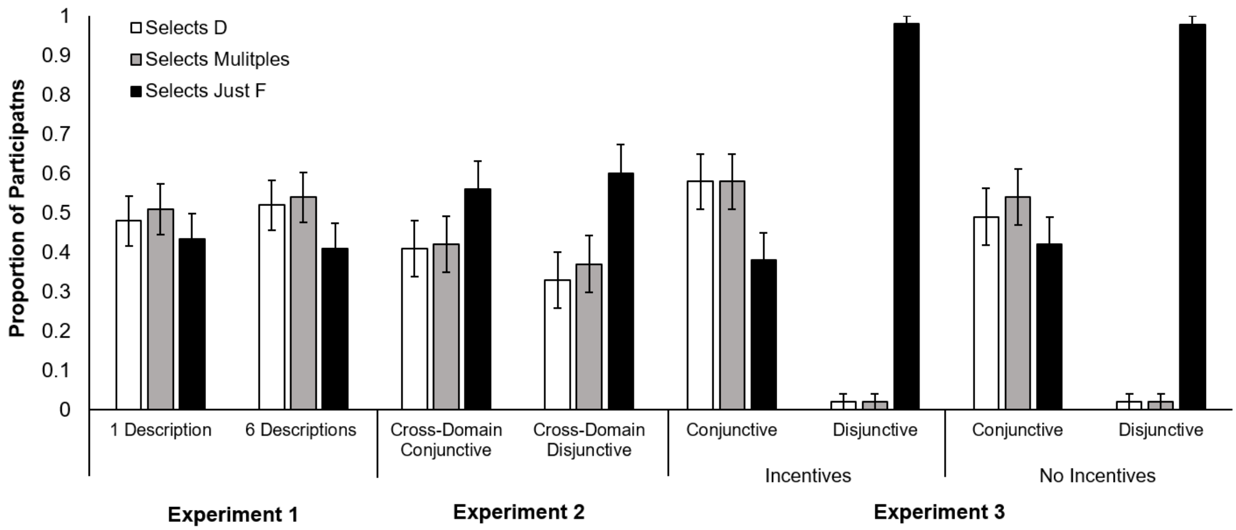

Figure 4.3 Average proportion of participants who selected object D, multiple objects, and just object F when asked to activate the machine in Experiments 1, 2, and 3. Error bars indicate one standard error of the mean in each direction.

Consistent with their blicket judgments, there were no significant differences in intervention choices between conditions. Adults in the 1 description condition were just as likely to select object D (*M* = .48; *SD* = .50), multiple objects (*M* = .51; *SD* = .50), and only object F (*M* = .43; *SD* = .50) as those in the 6 descriptions condition (object D: *M* = .52; *SD* = .50; multiple objects: *M* = .54; *SD* = .50; and only object F: *M* = .41; *SD* = .50), *p*'s > .10, Fisher's exact test.

4.2.2.3 Priming the causal learning task

To directly examine whether listing multiple descriptions encouraged adults to discover and endorse the conjunctive principle, we performed several binary logistic regressions with the condition, total number of priming responses, average novelty rating, and average appropriateness rating as predictors of blicket judgements and intervention choices. Our first model, which predicted whether participants judged D as a blicket, was not statistically significant $\chi^2(4) = 3.03$, $p = .55$, Nagelkerke $R^2 = .03$, suggesting that listing multiple descriptions failed to predict whether participants endorsed the conjunctively active D.

The last two models applied the same predictors as the first to examine the effect of the priming task on participants' intervention choices. The model to predict multiple objects was not statistically significant, $\chi^2(4) = 2.24$, $p = .692$, Nagelkerke $R^2 = .024$, nor was the model to predict just F, $\chi^2(4) = 6.20$, $p = .19$, Nagelkerke $R^2 = .07$.

4.2.3 Discussion

Participants tended to infer the more common disjunctive rule even when the evidence supported the more unusual conjunctive rule regardless of whether they were asked to list one or six descriptions. There are at least two possible explanations. One possibility is that generating descriptions may be a divergent thinking task even for those who were asked to list one description. Unlike convergent thinking which, as the name implies, has a single converging answer, divergent thinking has several possible responses. Since we did not specifically ask for the conventional description, it is possible that those who were asked to generate a single description were actually engaging in divergent thinking, selecting one from several possible descriptions. As a result, adults who listed one description for each object may then have been equally primed as those who listed six. Of note, however, is that adults who were instructed to list descriptions were just as likely to endorse the disjunctive rule as those who, as reported in Lucas et al. (2014), did not, thereby supporting an alternative interpretation that listing descriptions had little effect in encouraging adults to reason conjunctively.

A second possibility is that the unambiguous evidence in the two training trials was insufficient to encourage adults to overturn their strong disjunctive biases. Given that the conjunctive and disjunctive rules are abstract causal principles, or overhypotheses (Goodman, 1955), that apply across multiple contexts and domains, it is possible that the training trials, which only provide evidence in a single context, was inadequate to encourage a shift in adults' abstract higher-order reasoning. It may be, for instance, that adults require multiple examples in several different contexts that individually support the conjunctive rule and that collectively encourage the endorsement of this unusual abstract principle (see Gick & Holyoak, 1983). In Experiment 2, we explore this possibility by presenting adults with unambiguous cross-domain evidence across three training trials and examining whether they then infer the conjunctive rule.

4.3 Experiment 2

In Experiment 1, adults were asked to infer a causal relation within a single domain and context. In Experiment 2, we asked whether adults can infer this unusual conjunctive rule if they observed the rule applied across multiple domains. Adults observed cross-domain evidence that supported either the conjunctive or disjunctive rule, and afterwards, observed a toy activate in conjunction with multiple blocks. Similar to Experiment 1, adults then saw an ambiguous test trial that was consistent with both rules and were asked to demonstrate their abstract causal reasoning. If cross-domain evidence encourages adults to reason at the level of the overhypothesis, then adults who observe cross-domain conjunctive evidence should be more likely to endorse the conjunctive rule than those who observe cross-domain disjunctive evidence.

4.3.1 Method

4.3.1.1 Participants

One hundred undergraduate students, who were recruited from the same population as Experiment 1, participated in exchange for course credit (M_{age} = 22.39 years, SD_{age} = 5.67 years, 75 females). Participants were randomly assigned to one of two cross-domain conditions: the conjunctive condition (N = 50, M_{age} = 21.56 years, SD_{age} = 2.55 years, 41 females[1]) or the disjunctive condition (N = 50, M_{age} = 23.23 years, SD_{age} = 7.55 years, 34 females).

4.3.1.2 Materials

Materials were divided into three sets, one for each domain. The first set, which provided evidence in the physical domain, consisted of eight electrical components that connected together to create a circuit. A toy light turned on when the circuit was complete. The second set was modelled after Experiment 3 of Lucas et al. (2014) and provided evidence within the biological domain, specifically that some flowers make a toy monkey sneeze. Included in this set were eight plastic flowers, each of a different color, and a toy monkey. The final set was identical Experiment 1 and provided evidence in the physical domain. This set consisted of eleven blocks, each of a different shape, and a small toy hiding a wireless doorbell. Similar to the toy used in Experiment 1, this toy appeared to play music when objects were placed on top but was in fact surreptitiously operated by a switch controlled by the experimenter. Paper packets were provided for participants to record their responses.

4.3.1.3 Design

Participants were randomly assigned to one of two cross-domain conditions: the conjunctive condition or the disjunctive condition. In the conjunctive condition, participants observed two training trials, one in the physical domain with electrical circuits and the other in the biological domain with the flowers, that both supported the conjunctive rule. Those in the disjunctive condition, in contrast, observed cross-domain evidence supporting the disjunctive rule. The third training trial always, regardless of condition, provided evidence for the conjunctive rule with the musical toy. After these training trials, participants observed an ambiguous test trial with a new set of blocks that was consistent with both rules and were asked to demonstrate their causal reasoning. The order of the first two training trials was counterbalanced such that half of the participants started with the flowers, whereas the other half started with the electrical circuits. The third training trial as well as the final test trial were always with the musical toy and the blocks.

4.3.1.4 Procedure

Participants were tested in groups of no more than four. The procedure employed in Experiment 2 was identical to that of the causal learning task described in Experiment 1 with two exceptions: (1) participants observed three training trials across several domains rather than two training trials in a single domain, and (2) participants were told that objects producing a given effect were rare. Similar to Experiment 1, participants wrote their responses in paper packets.

Each training trial began with the experimenter presenting a container of objects and inviting participants to identify which objects were responsible for producing a particular effect. For example, in the biological training trial the experimenter placed a vase with three colorful flowers next to a toy monkey and invited participants to identify which flowers were *tulver flowers*. Participants were then told that tulver flowers cannot be identified by their appearance but have *tulverness* inside of them, which makes Monkey sneeze. To prevent Monkey from sneezing, participants were asked to help determine which ones were tulver flowers.

The experimenter then demonstrated that tulver flowers were rare by placing two pre-sorted vases on the table, one labeled as “Tulver Flowers” with a single flower and the other labelled as “Not Tulver Flowers” with four flowers. She then counted the number of flowers in each vase and commented that only one flower was a tulver flower, whereas the other four were not. After removing the pre-sorted vases, she placed the three unsorted flowers in a row on the table. The flowers as well as their left-right placement were counterbalanced across participants.

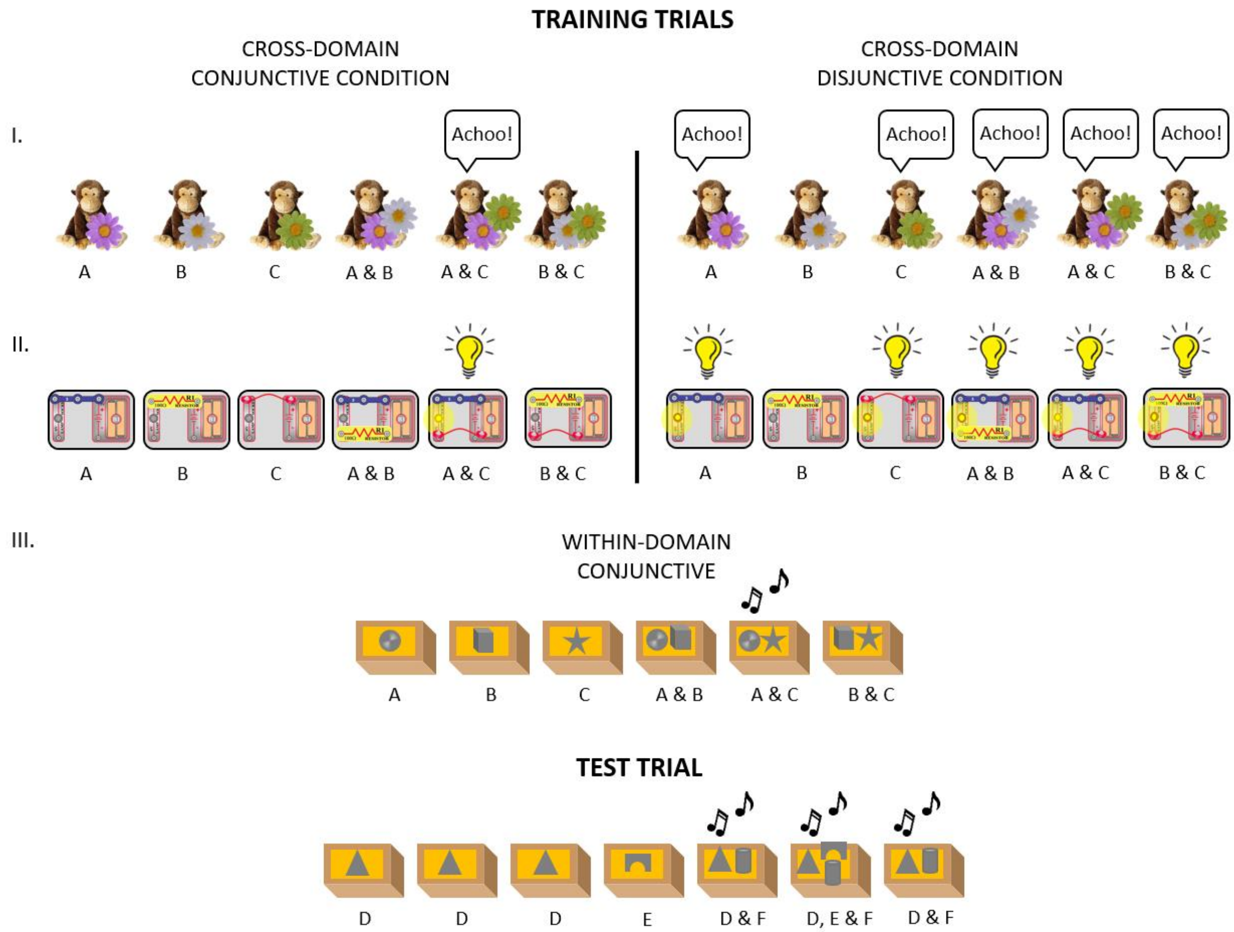

Figure 4.4 Activation patterns in the cross-domain conjunctive and disjunctive condition in Experiment 2.

After labelling each flower by its color, the experimenter demonstrated a series of events that resulted in Monkey either sneezing or not sneezing depending on the condition (see Figure 4.4). Those in the conjunctive condition, for example, observed that the combination of two specific flowers caused Monkey to sneeze. Those in the disjunctive condition, in contrast, observed individual flowers making Monkey sneeze. Participants were then asked whether or not each flower was a tulver flower. Once they recorded their responses, the experimenter removed Monkey and the three flowers and administered the next training trial.

The second training trial was identical in design to the first with the exception that participants were asked to identify which electrical components were *daxes*. Daxes cannot be identified by their appearance but have *daxiness* inside of them that, in this case, turn on a light. The experimenter then highlighted that daxes were rare by presenting two pre-sorted containers, one that was labelled as "Daxes" with one object and the other that was labelled as "Not Daxes" with four objects. Next, participants observed a series of events with three new components that supported either the conjunctive or disjunctive principle (see Figure 4.4). This was then followed by the experimenter asking whether or not each component was a dax. After participants recorded their responses, the experimenter removed the items and immediately administered the final training trial.

The third training trial was similar in design to the first two, but participants were shown a container of blocks and were asked to identify which ones were *blickets*. Participants were told that blickets cannot be identified by their appearance but have *blicketness* inside of them that, when placed on top of a novel toy, make the toy play music. After showing participants two pre-sorted containers--one with one blicket and the other with four non-blickets--the experimenter demonstrated a series of events with three new blocks that supported the conjunctive principle (see Figure 4.4). Afterwards, the experimenter asked whether each block was a blicket. At the end of the training trial, the experimenter immediately proceeded to the test trial, using the same procedure as described in Experiment 1 with three new blocks.

4.3.2 Results

To test for the effects of cross-domain evidence, we first analyzed participants' judgments on the three unambiguous training trials. We then examined whether adults endorsed the conjunctive or disjunctive rule by analyzing which objects they judged as blickets on the ambiguous test trial. All analyses were two-tailed.

4.3.2.1 Training trials

Figure 4.5 depicts the proportion of participants who labelled each object as a tulver flower, a dax, and a blicket in the three training trials. Regardless of the condition, participants should judge that objects A and C are responsible for producing an effect (i.e., causing Monkey to sneeze, turning on a light, playing music). The difference, however, is that adults in the conjunctive condition should believe that A and C are causally dependent on each other, whereas those in the disjunctive condition should believe that A and C are causally independent. In both conditions, participants should judge that object B is not causally responsible.

For each training trial, participants in the *disjunctive* condition were more likely to judge objects A and C as tulver flowers (A_{tulver}: M = .94; SD = .24; C_{tulver}: M = .94; SD = .27), daxes (A_{dax}: M = .94; SD = .24; C_{dax}: M = .88; SD = .33), and blickets ($A_{blicket}$: M = .39; SD = .49; $C_{blicket}$: M = .42; SD = .50) than object B (B_{tulver}: M = .08; SD = .27; B_{dax}: M = .06; SD = .24; $B_{blicket}$: M = .02; SD = .14), ps < .001, McNemar's test. There were no significant differences in their judgments of A and C, indicating that adults were equally likely to report that A was, for example, a dax as they were to report that C was a dax. Comparisons across training trials, however, yielded significant differences, with adults being less likely to judge objects A and C as blickets than they were to judge objects A and C as tulver flowers and daxes, ps < .001, McNemar's test. There were no reliable differences in their judgments of tulver flowers and daxes.

Participants in the *conjunctive* condition similarly judged objects A and C as tulvers (A_{tulver}: M = .43; SD = .50; C_{tulver}: M = .47; SD = .40), daxes (A_{dax}: M = .55; SD = .50; C_{dax}: M = .53; SD = .50), and blickets ($A_{blicket}$: M = .68; SD = .47; $C_{blicket}$: M = .64; SD = .49) more often than they judged object B as a tulver, a dax, and a blicket, respectively (B_{tulver}: M = .08; SD = .27; B_{dax}: M = .18; SD = .39; $B_{blicket}$: M = .06; SD = .24), ps < .01, McNemar's test. Adults did not differ in their judgments of A and C in each training trial. Unlike in the disjunctive condition, we found no reliable differences across trials with the exception that adults judged block A as a blicket more often than they judged flower A as a tulver flower, $p = .004$, McNemar's test.

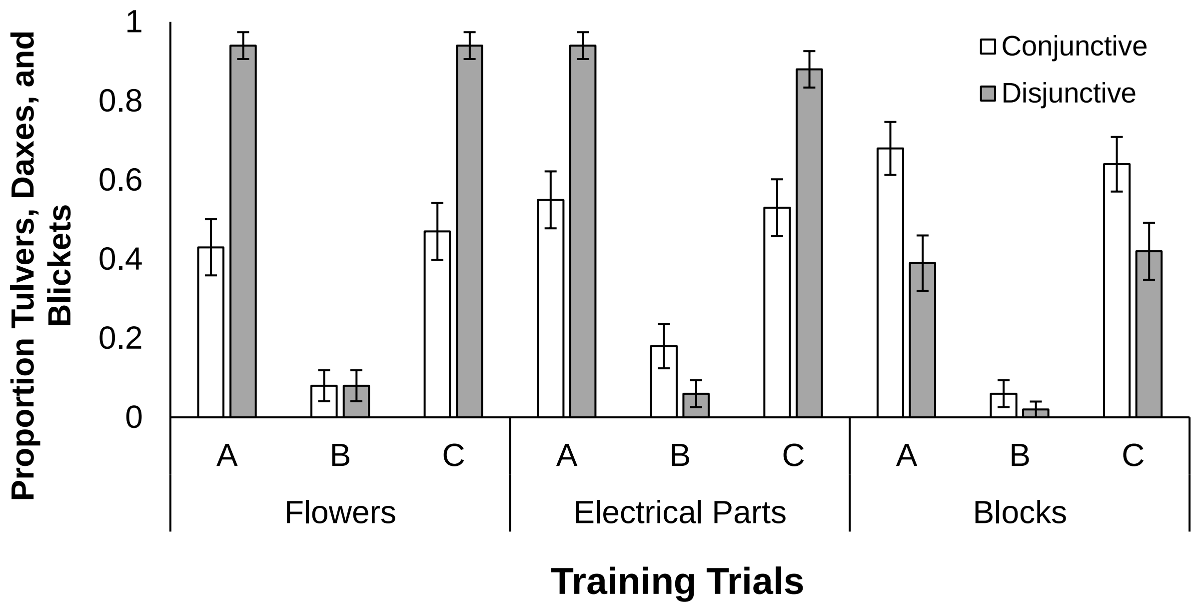

Figure 4.5 Average proportion of participants who judged each training object as a tulver, dax, and blicket in Experiment 2. Error bars indicate one standard error of the mean in each direction.

Comparisons between the two conditions reveal that adults in the disjunctive condition were more likely to judge objects A and C as tulver flowers and daxes than adults in the conjunctive condition, *p*s < .001, Fisher's exact test. They were, however, less likely than those in the conjunctive condition to judge blocks A and C as blickets, $p = .005$ and $p = .04$, respectively, Fisher's exact test, suggesting that those who observed two trials of cross-domain disjunctive evidence were, on the unambiguous conjunctive training trial, less likely to infer the conjunctive rule. There were no reliable differences in their judgments of object B across any of the three training trials.

4.3.2.1 Test trial

We next examined whether participants endorsed the conjunctive or disjunctive principle by analyzing their blicket judgments and intervention choices on the ambiguous test trial. If participants believed that combination of objects activated the machine, then they should label objects D and F as blickets and select both objects when asked to activate the machine. If, however, they believed that individual objects were responsible, then they should say judge object F as a blicket and only select that object during the intervention. Illustrated in Figure 4.2 is the mean proportion of blicket judgments for each object in the test trial. The mean proportion of intervention choices is shown in Figure 4.3.

Blicket Judgments. Consistent with our predictions, we found that adult assigned to the cross-domain *disjunctive condition* were more likely to label the unambiguous object F as a blicket (M = .80; SD = .41) than either the conjunctively active object D (M = .08; SD = .28), p < .001, McNemar's test, or the uncertain object E (M = .02; SD = .14), p < .001, McNemar's test, suggesting that adults inferred that the machine activated with a single block. There was no significant difference between adults' judgments of object D and E, p = .250, McNemar's test.

Those in the cross-domain *conjunctive condition* similarly judged object F to be a blicket (M = .92; SD = .27) more often than object D (M = .28; SD = .45), p < .001, McNemar's test, and object E (M = .08; SD = .27), p < .001, McNemar's test. However, unlike in the disjunctive condition, those in the conjunctive were also more likely to label object D as a blicket than object E, p = .002, McNemar's test.

When we compared performance across the two conditions, we found a significant difference in their judgment of block D, with those in the conjunctive condition reporting that block D was a blicket more often than those in the disjunctive condition, p = .018, Fisher's exact test. There were no significant differences in their judgments of E, p = .362, Fisher's exact test, or of F, p = .091, Fisher's exact test. Although this finding might suggest that adults who viewed cross-domain conjunctive evidence were more likely to endorse the conjunctive principle, post-hoc comparisons between adults who observed cross-domain training and a separate sample of adults from Lucas et al. (2014) who did not revealed a surprising effect. Participants who viewed cross-domain disjunctive evidence labelled block D as a blicket (M = .08; SD = .28) significantly less often than adults who only viewed conjunctive evidence in a single domain (M = .25; SD = .44), p = .017, Fisher's exact test. There were no reliable differences in the judgments of those who observed cross-domain conjunctive evidence (M = .28; SD = .45) from those who observed within-domain conjunctive evidence, p = .703, Fisher's exact test, indicating that the cross-domain training trials primed adults to reason more disjunctively but not more conjunctively.

Intervention Choices. Participants in the conjunctive condition were equally likely to select D (M = .41; SD = .50), include multiple objects (M = .42; SD = .50), and select just F (M = .56; SD = .50) compared to those in the disjunctive condition (object D: M = .33; SD = .48; multiple objects: M = .37; SD = .49; just F: M = .60; SD = .50), ps > .05, Fisher's exact test, suggesting those who viewed cross-domain conjunctive evidence were just as likely to endorse the conjunctive rule as were those who viewed cross-domain disjunctive evidence.

4.3.3 Discussion

The results from Experiment 2 illustrate the limitations of cross-domain evidence to elicit changes in adults' abstract causal reasoning. Adults who observed clear conjunctive evidence across domains largely inferred the disjunctive rule when the evidence was ambiguous. Indeed, adults were no more likely to infer the conjunctive rule after three cross-domain training trials than they were after two single-domain training trials. Performance was similar for those who observed cross-domain disjunctive evidence. The difference, however, was that these adults were even more likely to infer the disjunctive, suggesting that the disjunctive cross-domain evidence actually elicited even stronger disjunctive reasoning.

Unexpectedly, cross-domain evidence encouraged adults to reason disjunctively but not conjunctively. There are several reasonable explanations, one being that adults may have been victim to a confirmation bias and, as a result, were more sensitive to the familiar disjunctive evidence than the uncommon conjunctive evidence. Another possible explanation is that adults may have fallen into a learning trap by erroneously generalizing their past experience, which favored the disjunctive rule, to new ones (Denrell & March, 2001; Rich & Gurekis, 2018). This generalization, although adaptive when the environment is stable, can lead to premature exploitation of the disjunctive rule, thereby limiting whether adults learn the conjunctive rule. Finally, a third explanation is that adults may require even more evidence, both in quantity and in diversity, to overturn their strong bias for the disjunctive. That is, the belief that an individual is solely responsible for producing an effect may be so entrenched because of their prior experience that they may need more than three cross-domain training trials to think otherwise. These findings suggest that it is easier to prime strong prior beliefs than it is to prime weak ones.

It is clear from Experiments 1 and 2 that adults have difficulty inferring unusual hypotheses when they are primed to reason more broadly, opting instead to endorse familiar, yet at times incorrect, hypotheses. It is plausible that adults, when confronted with unfamiliar situations, adopt a strategy that exploits their previous experiences and their prior knowledge rather than one that encourages exploration and new discoveries. It seems reasonable then that adults endorse the more common disjunctive rule, for the disjunctive principle occurs with greater frequency than the conjunctive. However, as is the case in these experiments, the unusual hypothesis is sometimes a better explanation of the data. Can adults be encouraged to explore and discover new alternatives?

In Experiment 3, we tested the tradeoff between explore and exploit by presenting performance-based rewards. Unlike our previous experiments, performance-based monetary incentives provide a clearer and perhaps more effective test of exploration versus exploitation, particularly because agents must decide whether to exploit a well-known yet suboptimal belief—thereby leading to a smaller reward—or risk exploring other alternatives for the potential discovery of a more rewarding one. It is possible, for instance, that with incentives adults will show more sensitivity to the conjunctive data, recognize that their beliefs cannot explain their observations, and then explore, discover, and endorse the more unusual alternative. It is also possible, however, that performance-based incentives will encourage the repetition and exploitation of previously successful strategies and beliefs, thereby reducing flexibility and undermining performance. In Experiment 3, we test these two possibilities by examining how adults reason with conjunctive evidence when their performance is directly linked to monetary rewards.

4.4 Experiment 3

Similar to Experiments 1 and 2, adults observed a novel toy activate according to either the common disjunctive belief or the uncommon conjunctive belief and, afterwards, were asked to infer its causal structure. Unlike the previous experiments, however, adults were rewarded monetarily for their performance. If incentives increase adults' sensitivity to the evidence and encourage exploration, then those assigned to the conjunctive condition should endorse the conjunctive rule, stating that only specific combinations activate the toy. If, however, incentives encourage exploitation, then participants should endorse the disjunctive rule and claim that an individual object is responsible for activating the toy despite observing evidence suggesting otherwise. Those in the disjunctive condition, in contrast, should infer the disjunctive rule regardless of whether participants were incentivized to explore or exploit.

4.4.1 Method

4.4.1.1 Participants

Two hundred students and community members from a large university participated for monetary compensation (M_{age} = 21.60, SD_{age} = 4.00 years, 152 females). [2] Although the experiment was initially conducted in-person ($N_{in\text{-}person}$ = 75, M_{age} = 21.22, SD_{age} = 4.08 years, 53 females), the sudden emergence of a pandemic necessitated that the remaining participants be tested virtually ($N_{virtual}$ = 125, M_{age} = 21.83, SD_{age} = 3.96 years, 99 females). Twelve additional participants -- of which half were tested in-person and half were tested virtually -- were excluded due to experimenter error (8), equipment malfunction (3), or poor audio and video quality (1).

4.4.1.2 Materials

The stimulus materials were similar to those used in Experiments 1 and 2 and consisted of a toy machine that appeared to play music when blocks were placed on top. The toy, however, was activated by the experimenter with a hidden remote. There were also nine blocks, each with a different shape, and a small container holding the blocks.

4.4.1.3 Design

Participants were randomly assigned to either the conjunctive or the disjunctive condition and, within each of these conditions, were further assigned to either the incentive or no incentive condition. This combination yielded a total of four conditions with 50 participants each. Participants observed two training trials that supported either the conjunctive or disjunctive rule and then observed an ambiguous test trial that was consistent with both rules. Unlike the previous experiments, however, participants were instructed that they would be compensated and would automatically earn $5 for participating. Those who were also assigned to the incentive condition were then told that they could double their earnings, receiving an extra $5 if at the end of the experiment they correctly inferred the causal structure. Those in the no incentive condition were told that they would double their earnings regardless of performance. Participants tested in-person received cash earnings, whereas participants tested virtually received an electronic gift card.

4.4.1.4 Procedure

Participants were tested individually either in-person or virtually via Zoom. The experiment proper was identical for both testing methods with two exceptions. First, those who participated in-person received cash earnings immediately after the experiment, whereas those who participated virtually received an electronic gift card one to two weeks after participating. Second, virtual participants were instructed at the beginning of the experiment to use a desktop or laptop computer and afterwards were asked a series of follow-up questions to measure their attention during the experiment and to rate the quality of the video and audio. One participant rated the video quality as very poor and was excluded from the experiment.

The experimenter began by presenting a blank envelope and explaining to participants that this envelope was theirs. She then asked participants to write their name on the front or, for those who participated virtually, to spell out their names as she wrote on the envelope. The experimenter then placed the envelope aside. Next, she presented a container with nine blocks and explained that the goal is to identify which of the blocks are *blickets*. She then presented a novel toy and explained that, although blickets cannot be identified by their appearance, blickets have *blicketness* inside of them that when placed on top of the toy will make the toy play music.

The experimenter then presented \$5 as either cash or gift card and explained that participants would automatically receive \$5 for participating. She then placed the earnings in their envelopes. Next, she presented an additional \$5 and, for those in the incentive condition, told participants that they could double their earnings if they correctly identified which objects were blickets. Those in the no incentive condition, in contrast, were told that they would double their earnings regardless of performance. The experimenter then placed the second \$5 in the envelope for those in the no incentive condition.

After the experimenter explained the incentives, she selected three blocks, placing them in a row in front of the toy. Similar to Experiments 1 and 2, she then placed the blocks on top of the toy both individually and in combination with each other. The activation pattern differed by condition such that the toy activated according to either the conjunctive or the disjunctive rule. At the end of the training trial, participants were asked whether each object was a blicket. The experimenter then repeated the training trial with three new blocks.

Immediately after the second training trial, the experimenter proceeded to the test trial. Unlike Experiments 1 and 2, this test trial began with the experimenter reminding participants that they could double their earnings if they correctly identified which objects were blickets (incentive condition) or simply reminding them that they already doubled their earnings (no incentive condition). Then, similar to Experiments 1 and 2, participants observed an ambiguous activation pattern that was consistent with both the conjunctive and disjunctive rule. Afterwards, they were asked to judge whether each block was a blicket and to select those that would activate the toy.

At the end of the experiment, the experimenter removed the toy and the blocks and, for those who were tested virtually, asked three follow-up questions designed to measure their attention: (1) “What color were the blocks?”, (2) “What were the names of the last three objects?”, and (3) “What block did we try three times by itself?” She then asked participants to rate the quality of the audio and video on a scale from 1 to 5, with one being very poor and five being very good.

4.4.2 Results

Similar to Experiments 1 and 2, we examined participants’ blicket judgments and intervention choices on the ambiguous test trial. If participants endorse the disjunctive rule, then they should judge a single block, specifically block F, as a blicket, and they should only select this block when asked to activate the toy. Those who endorse the conjunctive rule, however, should judge and select a combination of blocks, such as blocks D and F. Unlike the previous experiments, we also tested for the unique effects of performance-based incentives. All analyses are two-tailed.

Blicket Judgments. Figure 4.2 illustrates the mean proportion of blicket judgments separated by condition. To test the effect of incentives and training evidence on blicket judgments, we performed a binary logistic regression with the causal training evidence, the incentive format, and the interaction between both of these variables as predictors of their judgment of block D. The model was statistically significant, $\chi^2(3) = 67.903$, $p < .001$, Nagelkerke $R^2 = .452$. Post-hoc comparisons yielded a main effect of training evidence, $\chi^2(1) = 51.572$, $p < .001$, with those in the conjunctive condition reporting D as a blicket ($M = .41$; $SD = .49$) more often than those in the disjunctive condition ($M = .00$; $SD = .00$). There was no main effect of incentive, $\chi^2(1) = 0.00$, $p = 1.00$, nor was there a significant interaction, $\chi^2(1) = 0.00$, $p = 1.00$.

To closely examine the effect of the training evidence, we collapsed across the incentive groups and analyzed the conjunctive and disjunctive conditions separately. Participants in the *disjunctive condition* reported block F as a blicket ($M = .99$; $SD = .10$) more often than block D ($M = .00$; $SD = .00$), $p < .001$, McNemar's test, or block E ($M = .05$; $SD = .22$), $p < .001$, McNemar's test. There was no significant difference in their judgments of D and E, $p = .063$, McNemar's test. Those in the *conjunctive condition* similarly labelled F as a blicket ($M = .80$; $SD = .40$) significantly more often than they labelled D ($M = .41$; $SD = .49$), $p < .001$, McNemar's test, or E as blickets ($M = .21$; $SD = .41$), $p < .001$, McNemar's test. Unlike in the disjunctive condition, participants in the conjunctive condition were more likely to say that D was a blicket than E, $p < .001$, McNemar's test.

We then performed two secondary analyses to predict whether participants labelled E and F as blickets using the same predictors. Once again, our binary logistic regressions reliably predicted whether participants reported E, $\chi^2(3) = 14.143$, $p =.003$, Nagelkerke $R^2 = .127$, and F as blickets, $\chi^2(3) = 27.739$, $p < .001$, Nagelkerke $R^2 = .238$. Although there was no effect of incentives nor was there a significant interaction, *p*s > .10, there was a main effect of training evidence on E, $\chi^2(1) = 6.066$, $p = .014$, and F, $\chi^2(1) = 19.207$, $p < .001$. Those in the conjunctive condition were more likely to report E as a blicket but were less likely to report F as a blicket compared to those in the disjunctive condition.

Intervention Choices. The mean proportion of intervention choices is shown in Figure 4.3. Similar to Experiments 1 and 2, we performed several analyses to examine the effect of the training evidence and incentives on participants' interventions. Using a logistic regression with the type of evidence, the use of incentives, and the interaction between these variables as predictors, we found the model reliably predicted whether participants selected the conjunctive active D, $\chi^2(3) = 78.469$, $p < .001$, Nagelkerke $R^2 = .472$. Post-hoc comparisons showed a significant effect of the conjunctive and disjunctive evidence, with participants in the conjunctive condition selecting block D ($M = .54$; $SD = .50$) at a greater rate than those in the disjunctive condition ($M = .02$; $SD = .14$), $\chi^2(1) = 16.111$, $p < .001$. There was no main effect of incentive, $\chi^2(1) = 0.00$, $p = .988$, nor was there an interaction between the type of training evidence and the incentive, $\chi^2(1) = .067$, $p = .796$.

We performed a similar logistic regression to predict whether participants selected multiple objects and found a similar pattern. This model was statistically significant, $\chi^2(3) = 83.583$, $p < .001$, Nagelkerke $R^2 = .489$, with those in the conjunctive condition selecting multiple objects ($M = .56$; $SD = .50$) more often than those in the disjunctive condition ($M = .02$; $SD = .14$), $\chi^2(1) = 16.111$ $p < .001$. There was no effect of incentive, $\chi^2(1) = 0.00$, $p = .988$, and there was no reliable interaction, $\chi^2(1) = .015$, $p = .902$.

A third logistic regression was performed to predict whether participants selected just block F, thereby endorsing the disjunctive rule. Consistent with the first two analyses, the model was statistically significant, $\chi^2(3) = 92.893$, $p < .001$, Nagelkerke $R^2 = .525$, with a main effect of training evidence, $\chi^2(1) = 17.368$, $p < .001$. Those in the disjunctive condition were more likely to select just F ($M = .98$; $SD = .14$) than those in the conjunctive condition ($M = .40$; $SD = .49$).

Once again, there was no effect of incentives, $\chi^2(1) = 0.00$, $p = .988$, nor was there an interaction, $\chi^2(1) = .016$, $p = .900$.

Effect of financial compensation. Although our results might suggest that adults who observed conjunctive evidence responded conjunctively whereas those who observed disjunctive evidence responded disjunctively, we examined the role of financial compensation independent of performance-based incentives and discovered an interesting effect. To explore how money shifted performance, we compared adults who, in Experiment 3, received payment for participating against those who, in Lucas et al. (2014), received course credit. If adults become more sensitive to the data simply because they are monetarily compensated (see Camerer & Hogarth, 1999), then adults in Experiment 3 should be more sensitive to the evidence than those in Lucas et al. Collapsing across the incentive and no incentive group, we found no significant difference in the conjunctive condition between participants who received financial payment and those who received course credit. They were just as likely to report D ($M_{Lucas\ et\ al.} = .25$; $SD_{Lucas\ et\ al.} = .44$), E ($M_{Lucas\ et\ al.} = .11$; $SD_{Lucas\ et\ al.} = .32$), and F as blickets ($M_{Lucas\ et\ al.} = .71$; $SD_{Lucas\ et\ al.} = .46$) as well as to include D ($M_{Lucas\ et\ al.} = .35$; $SD_{Lucas\ et\ al.} = .49$), select multiple objects ($M_{Lucas\ et\ al.} = .36$; $SD_{Lucas\ et\ al.} = .49$), and select only F ($M_{Lucas\ et\ al.} = .57$; $SD_{Lucas\ et\ al.} = .50$), *p*s > .05, Fisher's exact test, suggesting that financial compensation did not increase adults' sensitivity to the conjunctive data.

There were, however, significant differences in the disjunctive condition. Participants who received payment became even more disjunctive in their response, reporting that F was a blicket with greater frequency ($M_{Lucas\ et\ al.} = .82$; $SD_{Lucas\ et\ al.} = .39$), $p = .002$, Fisher's exact test, and D with less frequency ($M_{Lucas\ et\ al.} = .11$; $SD_{Lucas\ et\ al.} = .32$), $p = .010$, Fisher's exact test. There was no significant difference in their judgment of E ($M_{Lucas\ et\ al.} = .11$; $SD_{Lucas\ et\ al.} = .32$). Their intervention choices reflected a similar pattern. Participants in Experiment 3 were more likely to select just F ($M_{Lucas\ et\ al.} = .79$; $SD_{Lucas\ et\ al.} = .42$), $p = .001$, Fisher's exact test, and hence were less likely to select multiple blocks ($M_{Lucas\ et\ al.} = .15$; $SD_{Lucas\ et\ al.} = .36$), $p = .018$, Fisher's exact test, compared with those who received course credit. They were equally likely to select D in both experiments ($M_{Lucas\ et\ al.} = .08$; $SD_{Lucas\ et\ al.} = .27$). Taken together, these findings indicate that the difference we observed between the conjunctive and disjunctive conditions in Experiment 3 was driven by a heightened sensitivity to the disjunctive evidence rather than an increased sensitivity to evidence more generally.

4.4.3 Discussion

Contrary to our prediction, incentives failed to elicit exploration or exploitation. Adults whose earnings depended on their performance were just as likely to endorse either rule as those whose earnings were not performance-based. Although adults in the conjunctive condition responded more conjunctively in both judgments and intervention choices than those in the disjunctive condition, this effect was largely driven by an increased sensitivity to disjunctive evidence. Indeed, adults in both the conjunctive and disjunctive condition continued to respond disjunctively, reporting that block F was a blicket more often than blocks D or E. The difference, however, was that adults in the disjunctive condition were more likely to infer a single block as being causally responsible if they received monetary compensation instead of course credit.

Our finding that performance-based incentives failed to elicit conjunctive or disjunctive reasoning is not entirely surprising given that the effects of incentives are often mixed (see Camerer & Hogarth, 1999). Incentives can improve performance just as it can, in some cases, hurt. Most commonly, however, incentives appear to have little to no effect on mean performance. Such may be the case if the task is either too hard or too easy, resulting in either floor or ceiling effects. With this interpretation, it is possible that adults in the disjunctive condition perceived the task as being too easy, whereas adults in the conjunctive condition perceived it as too hard. It is worth noting that although incentives did not affect mean performance, it did reduce variation in the disjunctive condition perhaps by reducing unmotivated outliers. Adults in the disjunctive condition were even more likely to infer a single cause when they received money than when they received course credit. This difference may reflect a difference in extrinsic motivation, with financial rewards leading to increased effort, or it may simply reflect a difference in the sample. In particular, those who received money volunteered to participate, possibly demonstrating their intrinsic motivation, whereas those who received course credit did so out of obligation.

Although it is clear from our previous work that adults have difficulty inferring unusual hypotheses even when they are primed or incentivized to reason more broadly, it is possible that the manipulations were too subtle. How will adults respond if the cortical region that regulates thought and controls behavior is suppressed? Recent work using transcranial direct current stimulation (tDCS)--a noninvasive technique that can either increase (anodal stimulation) or decrease (cathodal stimulation) cortical activity--has demonstrated that reducing excitability in the left prefrontal cortex stimulates creative insight by lowering cognitive control (Chi & Snyder, 2011, 2012; Chrysikou et al., 2013, Kounios & Beeman, 2014). In Experiment 4, we use tDCS to apply cathodal stimulation to suppress activity in the prefrontal cortex as adults observe conjunctive or disjunctive evidence.

4.5 General Discussion

Across three experiments, we demonstrate a striking shift in the inflexibility of adults' beliefs. In particular, we found that adults continue to endorse a prior belief that is weakly supported by the data even after being primed or incentivized to think creatively and broadly. Irrespective of whether adults listed one or six descriptions, they tended to infer the more common disjunctive rule despite observing evidence that supported the conjunctive (Experiment 1). Even when they observed the conjunctive rule applied across multiple domains and contexts, adults continued to believe that a single object was responsible for producing an effect (Experiment 2). Adults who were incentivized to respond correctly performed no differently. Despite observing evidence that supported the conjunctive rule, they still endorsed the disjunctive (Experiment 3). Taken together these results, which are consistent with previous work (Lucas et al., 2014; Wente et al., 2019), demonstrate the remarkable rigidity as well as robustness of adults' beliefs.

Although children can easily overturn their bias for the disjunctive rule when shown evidence supporting the conjunctive, it is unclear why adults, who have more cognitive resources than children, have such difficulty revising their beliefs. Do adults fail to endorse the conjunctive because they fail to generate the competing hypothesis, or do they generate the competing hypothesis but then fail to evaluate this against the data? As mentioned above, the ability to generate hypotheses is not mutually exclusive from the weight of the prior. In fact, those with stronger priors should be less likely to consider other alternatives, whereas those with weaker priors should be more likely to search through their hypothesis space. Given that adults have more experience interacting with the world than children, they necessarily have stronger priors, especially for the disjunctive rule, and therefore may be less likely to generate an alternative hypothesis. As a result, adults may never even consider the conjunctive rule as a possibility. In order to encourage adults to search, it may be necessary then to present even more counterevidence against the disjunctive to decrease the likelihood that it is true. Since we could neither control nor measure the weight of adults' prior beliefs to that of children's, it is possible that the difference in performance is largely due to a difference in their priors (and hence their search) rather than a difference in their evaluation of the data. Indeed, adults can evaluate the conjunctive evidence if the toy is designed to have two receiving slots, thereby implying that two blocks, one for each slot, are needed to activate the toy. Without such design cues to encourage the conjunctive hypothesis, however, adults have difficulty generating this more unusual alternative (Walker, Rett, & Bonawitz, 2020).

The findings from these three experiments suggest that our manipulations were insufficient to increase adults' sensitivity to evidence more generally. Instead, our manipulations had a more targeted effect when the evidence was consistent with their biases and beliefs. Adults were even more likely to infer a single cause if they had observed disjunctive evidence across multiple domains or if they had been compensated monetarily. When the evidence was inconsistent with their beliefs, however, our manipulations were insufficient to promote wide searches that would facilitate conjunctive reasoning. This distinction may demonstrate that adults were subject to a confirmation bias, or it may be that they fell into a learning trap (e.g., Rich & Guerckis, 2018). In both instances, however, adults are victims of their own knowledge.

This work demonstrates that adults, despite being primed, incentivized, or stimulated, have difficulty overturning their strong priors. This is particularly evident when the evidence supports an unusual alternative, such as the conjunctive rule. This apparent insensitivity to the evidence might stem from a more general inflexibility to generate alternative hypotheses, or it might be because adults have difficulty evaluating the evidence. Although our findings cannot differentiate between these two possibilities, we do provide strong evidence demonstrating the rigidity and robustness of adults' beliefs. In doing so, we lend some truth to the adage that you cannot teach an old dog new tricks (or, at the very least, it is challenging). But with enough training and evidence perhaps we, as adults, can learn a new thing or two. That, however, is for future work.

Chapter 5

Conclusions

5.1 Conclusion and implications of the empirical work

Our experiences are often noisy, sparse, and ambiguous. Oftentimes this messiness is expected variance that has little consequence on our interpretation of the data. Indeed, we quickly and oftentimes accurately draw inferences from these less-than-ideal experiences. Sometimes, however, this messiness captures qualitatively different experiences. How, then, do we identify data that reflect meaningful change from those that reflect trivial noise?

In this dissertation, I examined the factors that encourage as well as discourage belief revision in children and adults. Chapter 2 showed that 4- and 5-year-olds behave rationally when revising their higher-order beliefs, updating their beliefs when the counterevidence strongly outweighed the prior but maintaining their beliefs when it did not. Even younger children are rational revisionists, as shown in Chapter 3. This ability to flexibly change beliefs in response to new information is independent of the top-down executive flexibility that controls attention and behaviors. Finally, Chapter 4 illustrates the robustness of adults' beliefs, demonstrating the many ways that adults do not revise their beliefs. Taken together, these empirical studies illustrate the remarkable developmental decline in epistemic flexibility, contrasting the ease at which children rationally update their beliefs against adults' impressive stubbornness to stick with theirs.

5.1.1 Children rationally revise their higher-order beliefs

Chapter 2 demonstrates that 4- and 5-year-olds not only revise their abstract, higher-order beliefs in response to new information but do so rationally. When shown a single event that violates a strongly supported belief, children largely maintain their initial higher-order belief and ignore or discount the counterevidence. If, however, they observe more evidence that conflicts with this belief, then they change their minds. This study, although simple in design, carries important implications in clarifying our understanding of how children respond to new information. In particular, our work demonstrates that young children reason in ways that are consistent with Bayesian learning at a level of abstraction that is even higher than previously demonstrated (Bonawitz et al., 2012; Schulz et al., 2007). From relatively little data, children can form and, as shown in Chapter 2, revise abstract, higher-order generalizations. It is impressive that even in the absence of direct experience, children make powerful inductive inferences that can be easily changed. What is altogether more impressive is that children are doing so rationally, weighing the strength of the supporting data against the strength of their prior knowledge.

5.1.2 Two forms of cognitive flexibility

Chapter 3 highlights a tension surrounding children's cognitive flexibility. In some contexts, young children are rational, flexible thinkers, computing statistical probabilities to update their beliefs as the data change (e.g., Bonawitz et al., 2012; Kushnir & Gopnik, 2007; Schulz et al., 2007). Yet in other contexts, they appear to be irrationally inflexible. Preschoolers can, for example, flawlessly sort bi-variant cards along a single dimension but, when asked to switch and sort by another dimension, they perseverate (e.g., Zelazo et al., 2003). Why do children appear cognitively flexible in some cases yet cognitively inflexible in others?

In Chapter 3, I address this apparent paradox by proposing two distinct forms of cognitive flexibility: (1) epistemic flexibility, which captures changes in knowledge due to the introduction of new evidence, and (2) executive flexibility, which captures deliberate shifts in goal-directed behaviors. Experiments 1 and 2 replicated our earlier finding that young children are rational revisionists, showing that even 3-year-olds normatively update their beliefs depending on the statistical regularities of the evidence. Notably, this ability to change beliefs in response to new evidence is independent of executive functions. Three-year-olds revised their higher-order beliefs regardless of whether they correctly switched when asked to shift from one sorting rule to another or incorrectly perseverated. This suggests that these two forms of flexibility are not only distinct but are also dissociable from one another. Experiment 3 provides additional support for this distinction by demonstrating that even in a non-causal context, 3-year-olds continued to make and update social beliefs in response to strong counterevidence.

This contrast between epistemic and executive flexibility carries important implications, particularly as it is consistent with a recent idea that proposes an age-related tradeoff between exploration and exploitation (Gopnik et al., 2017). Young children often learn by exploring and conducting play-based experiments to gather data about their environment (e.g., Bonawitz et al., 2012; Stahl & Feigenson, 2015). This heightened period of exploration leads children to discover new information but—as they grow older, acquire more knowledge, and strengthen their priors—eventually gives way to a strategy of exploitation that heavily recruits executive functions. These two learning strategies, which emerge at different stages in development, parallel the two forms of cognitive flexibility, with exploration relating to epistemic flexibility and exploitation relating to executive flexibility.

5.1.3 Adults' robust beliefs

Chapter 4 examines a developmental decline in adults' willingness to learn from data and revise their beliefs. It is rather surprising that adults are sometimes worse than children at considering new evidence, particularly since adults have more mature cognitive resources available (Lucas et al., 2014; Wente et al., 2017). One possible explanation for their poor performance is that adults may be less likely to discover other alternatives and update their beliefs because they hold strong biases that constrain the hypothesis space. In Chapter 4, I tested this possibility by trying to prime or incentivize adults to elicit exploratory behaviors. Instead, these three experiments illustrate the entrenchment of adults' beliefs and their remarkable resistance to change.

In Experiment 1, adults were instructed to generate 1 or 6 descriptions for common objects with the assumption that listing six descriptions required a broader, exploratory search than listing one description. Afterwards, adults observed an unusual toy machine activate when particular combinations of blocks were placed on top of it. Despite generating multiple descriptions and despite observing evidence that supported the unusual belief that specific combinations of blocks activated the toy whereas individual blocks did not (i.e., conjunctive rule), adults failed to override their assumption that a single block was causally responsible (i.e., disjunctive rule). That is, priming adults to generate multiple descriptions for each item failed to encourage adults to endorse the more unusual, yet more appropriate, conjunctive hypothesis. Experiment 2 then replicated these findings by demonstrating that adults continued to believe the disjunctive rule even after they observed evidence that unambiguously supported the conjunctive rule across multiple domains and contexts. There was, however, an effect of disjunctive evidence, suggesting that adults are easily persuaded when the evidence confirms rather than violates their prior beliefs. Even performance-based financial incentives failed to motivate exploration and elicit significant belief revision. In Experiment 3, adults whose earnings depended on their performance responded similarly to those whose earnings did not. Indeed, it appears that simply receiving payment increased adults' sensitivity to disjunctive but not conjunctive evidence. Similar to the first two experiments, adults continued to incorrectly believe that an individual block was responsible for activating a toy regardless of whether their earnings depended on their inferences.

Together, these findings demonstrate adults' astounding resistance to counterevidence. Despite our attempts to encourage epistemic flexibility, adults continued to endorse a belief that was largely unsupported by data. This failure to revise does not appear to be a problem of hypothesis evaluation particularly since adults accept the conjunctive rule if the toy is designed to encourage a conjunctive interpretation (Walker et al., 2020). Instead, our results may reflect a problem of hypothesis generation, of searching through the hypothesis space and arriving at the alternative possibility that specific combinations of objects activate the toy. In particular, it may be that adults' strong prior for the disjunctive rule precludes them from searching beyond what is already accepted. As mentioned earlier, the ability to generate alternative hypotheses is not mutually exclusive from the weight of the prior. In fact, those with stronger priors should be less likely to consider alternatives, whereas those with weaker priors should have flatter distributions and as a result should be more likely to search through their hypothesis space. Given that adults have years of experience interacting with the world, they necessarily have stronger priors and, as a result, could simply require more counterevidence to overturn their belief. Although it is challenging for adults to override their belief, it is notable that their priors can be strengthened as demonstrated by the ease at which adults became even more disjunctive when the evidence supported an individual cause. As hard as it is to revise beliefs, it is much easier to strengthen and entrench current beliefs.

5.2 Remaining questions and future directions

What is the structure of our belief system, and how do the different levels of abstraction interact to inform one another? The work presented in Chapters 2 and 3 provide evidence that children apply their revised higher-order beliefs to make inferences about unfamiliar lower-level instances. When asked, for example, to activate a new toy, they reliably selected the block that matched their higher-order belief. Note that children were always asked to generalize to a new toy. Although this design was necessary to measure changes in higher-order beliefs, it leaves open the question of whether children, in revising their higher-order belief, are also revising conflicting lower-order beliefs, namely their belief about the toy that was activated with a different rule. It is possible that children interpret this conflicting event as anomalous, thereby continuing to believe that this toy activates according to the initial rule. Indeed, children often explain away counterevidence to preserve their beliefs by appealing to auxiliary variables (Bonawitz et al., 2012) or hidden entities (Schulz, Goodman, Tenenbaum, & Jenkins, 2008). It is also possible, however, that when children revise their higher-order belief they are simultaneously updating and aligning all lower-order beliefs. Doing so, in fact, would simplify the belief system, making for an attractive explanation (Bonawitz & Lombrozo, 2012). In ongoing work, I am addressing these two possibilities by asking whether or not children update their conflicting lower-level beliefs to match their revised higher-order belief.

Belief revision is a complex and onerous process. It is not enough to simply detect conflicting evidence. Learners must also generate alternative hypotheses that can explain their observations, evaluate these competing hypotheses against the data, and then decide how to update their beliefs given the interaction between the strength of the new evidence and the weight of their prior beliefs. The results from Chapter 2 and 3 offer some insight into this process by demonstrating that children can track statistical probabilities to evaluate the new data against the prior. This strategy, however, is taxing and prone to mistakes. An easier, less costly alternative might be to rely on the knowledge of other individuals in the environment, such as the case when students learn from teachers. Indeed, there is considerable research showing that children identify knowledgeable informants from ignorant ones and then leverage that information to decide whom to trust (e.g., Koenig & Harris, 2005). Similar reasoning might then apply to learners who have conflicting knowledge. Relying on a reliable source might, in this case, bypass the need to generate alternative hypotheses and reduce the uncertainty of what information is valid. It is worth noting that social learning, although helpful in transmitting knowledge, is constrained by the teacher, especially since pedagogy greatly reduces exploration (Bonawitz et al., 2011). How social information interacts with our changing beliefs, although important in expanding our understanding of belief revision, is subject for further study.

In this dissertation, belief revision is characterized as changing from one belief to another, from believing, for example, that matching color blocks activate a set of toys to now believing that matching shape blocks do. But, belief revision does not need to be the complete replacement of a prior belief. Instead, a belief can also be updated through expansion or contraction (Thagard, 1992). That is, for instance, children could have revised their belief to one that was consistent with both rules, selecting both the color and the shape block. Indeed, revisions that incorporate the prior belief may be more common, especially if knowledge constrains the hypothesis search to alternatives that resemble the prior. We might then expect a developmental interaction in the type of belief revisions, with younger learners being more likely to make extreme revisions that disregard their prior beliefs and older ones making minor adjustments. In future work, it will be important to clarify the types of belief revision to better understand how knowledge is updated.

5.3 Concluding remarks

This dissertation shows how children and adults diverge in their response to evidence that violates their beliefs, demonstrating that younger learners are more willing to update their beliefs than older ones. In particular, preschoolers rationally revise their higher-order beliefs, whereas adults, who have more cognitive resources than children, fail even after observing counterevidence that strongly contradicts their beliefs. This heightened period of epistemic flexibility in childhood giving way to a notable decline in adulthood may reflect an adaptive learning strategy of exploration leading to exploitation.

Although this work describes seemingly trivial beliefs about whether certain blocks activate specific toys, this research has the power to clarify important issues. Why, for instance, are some people against vaccines despite their observed effectiveness in protecting against diseases? Why do others claim that climate change is a myth even though temperatures are rising and ice caps are shrinking? And why do people believe the misinformation that is spreading through social media? It is scary to think, and even more terrifying to witness, that our beliefs are so entrenched that they cannot be changed. But, they can. We can change our perspective to consider another's viewpoint, we can change our behavior to be more mindful of our actions, and we can change our beliefs to demand cultural, political, and social reforms. It is hard work, for sure. But others have done it before and so can we.

References

Amsterlaw, J., & Wellman, H. M. (2006). Theories of mind in transition: A microgenetic study of the development of false belief understanding. *Journal of Cognition and Development*, *7*, 139-172. https://doi.org/10.1207/s15327647jcd0702_1

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

Bonawitz, E., Fisher, A., & Schulz, L. (2012) Teaching three-and-a-half-year-olds to revise their beliefs given ambiguous evidence. *Journal of Cognition and Development*, *13*, 266-280. https://doi.org/10.1080/15248372.2011.577701

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

Bunge, S. A., & Zelazo, P. D. (2006). A brain-based account of the development of rule use in childhood. *Current Directions in Psychological Science*, *15*(3), 118-121. https://doi.org/10.1111/j.0963-7214.2006.00419.x

Camerer, C. F., & Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty*, *19*(1-3), 7-42. https://doi.org/10.1023/A:1007850605129

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology*, *28*(2), 595-616. https://doi.org/10.1207/s15326942dn2802_3

Chi, R. P., & Snyder, A. W. (2011). Facilitate insight by non-invasive brain stimulation. *PloS one*, *6*(2), e16655. https://doi.org/10.1371/journal.pone.0016655

Chi, R. P., & Snyder, A. W. (2012). Brain stimulation enables the solution of an inherently difficult problem. *Neuroscience Letters*, *515*(2), 121-124. https://doi.org/10.1016/j.neulet.2012.03.012

Christensen, P. R., & Guilford, J. P. (1958). Creativity/Fluency Scales. Beverly Hills, CA: Sheridan Supply.

Chrysikou, E. G. (2006). When shoes become hammers: Goal-derived categorization training enhances problem-solving performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(4), 935–942. https://doi.org/10.1037/0278-7393.32.4.935

Chrysikou, E. G., Hamilton, R. H., Coslett, H. B., Datta, A., Bikson, M., & Thompson-Schill, S. L. (2013). Noninvasive transcranial direct current stimulation over the left prefrontal cortex facilitates cognitive flexibility in tool use. *Cognitive Neuroscience*, *4*(2), 81-89. https://doi.org/10.1080/17588928.2013.768221

Denrell, J., & March, J. G. (2001). Adaptation as information restriction: The hot stove effect. *Organization Science*, *12*(5), 523-538. https://doi.org/10.1287/orsc.12.5.523.10092

Dewar, K. M., & Xu, F. (2010). Induction, overhypothesis, and the origin of abstract knowledge: Evidence from 9-month-old infants. *Psychological Science*, *21*, 1871-1877. https://doi.org/10.1177/0956797610388810

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

Diamond, A., Carlson, S. M., & Beck, D. M. (2005). Preschool children's performance in task switching on the dimensional change card sort task: Separating the dimensions aids the ability to switch. *Developmental neuropsychology*, *28*(2), 689-729. https://doi.org/10.1207/s15326942dn2802_7

Duncker, K. (1945). On problem-solving. *Psychological Monographs*, *58*(5, Whole No. 270).

Frith, C. D. (2000). The role of dorsolateral prefrontal cortex in the selection of action as revealed by functional imaging. *Control of cognitive processes: Attention and performance XVIII*, 549-565.

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin, 134*(1), 31–60. https://doi.org/10.1037/0033-2909.134.1.31

German, T. P., & Defeyter, M. A. (2000). Immunity to functional fixedness in young children. *Psychonomic Bulletin & Review*, *7*(4), 707-712. https://doi.org/10.3758/BF03213010

Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3 1/2-7 years old on a Stroop-like day-night test. *Cognition*, *53*(2), 129-153. https://doi.org/10.1016/0010-0277(94)90068-X

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*(1), 1-38. https://doi.org/10.1016/0010-0285(83)90002-6

Goddu, M., Lombrozo, T., & Gopnik, A. (2020). Transformations and transfer: Preschool children understand abstract relations and reason analogically in a causal task. *Child Development, 91*(6), 1898-1915. https://doi.org/10.1111/cdev.13412

Goodman, N. (1955). Fact, fiction, and forecast. Cambridge, MA: Harvard University Press.

Gopnik, A. (2020). Childhood as a solution to explore-exploit tensions. *Philosophical Transactions of the Royal Society*. https://doi.org/10.1098/rstb.2019.0502

Gopnik, A., Griffiths, T. L., & Lucas, C. G. (2015). When younger learners can be better (or at least more open-minded) than older ones. *Current Directions in Psychological Science, 24*(2), 87-92. https://doi.org/10.1177/0963721414556653

Gopnik, A., O'Grady, S., Lucas, C. G., Griffiths, T. L., Wente, A., Bridgers, S., Aboody, R., Fung, H., & Dahl, R. E. (2017). Changes in cognitive flexibility and hypothesis search across human life history from childhood to adolescence to adulthood. *Proceedings of the National Academy of Sciences*, *114*(30), 7892-7899. https://doi.org/10.1073/pnas.1700811114

Gopnik, A., & Sobel, D. M. (2000). Detecting blickets: How young children use information about novel causal powers in categorization and induction. *Child Development*, *71*, 1205–1222. https://doi.org/10.1111/1467-8624.00224

Hennessey, B. A., & Amabile, T. M. (2010). Creativity. *Annual Review of Psychology*, *61,* 569-598.

Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science, 10*, 307-321. https://doi.org/10.1111/j.1467-7687.2007.00585.x

Kharitonova, M., Chien, S., Colunga, E., & Munakata, Y. (2009). More than a matter of getting 'unstuck': Flexible thinkers use more abstract representations than perseverators. *Developmental Science*, *12*(4), 662-669. https://doi.org/10.1111/j.1467-7687.2008.00799.x

Kimura, K., & Gopnik, A. (2019). Rational Higher-Order Belief Revision in Young Children. *Child Development*, *90*(1), 91-97. https://doi.org/10.1111/cdev.13143

Kirkham, N. Z., Cruess, L., & Diamond, A. (2003). Helping children apply their knowledge to their behavior on a dimension-switching task. *Developmental Science, 6*(5), 449-467. https://doi.org/10.1111/1467-7687.00300

Kloo, D., & Perner, J. (2005). Disentangling dimensions in the dimensional change card-sorting task. *Developmental Science*, *8*(1), 44-56. https://doi.org/10.1111/j.1467-7687.2005.00392.x

Kounios, J., & Beeman, M. (2014). The cognitive neuroscience of insight. *Annual Review of Psychology, 65*, 71-93. https://doi.org/10.1146/annurev-psych-010213-115154

Kushnir, T., & Gopnik, A. (2007). Conditional probability versus spatial contiguity in causal learning: Preschoolers use new contingency evidence to overcome prior spatial assumptions. *Developmental Psychology*, *43*, 186-196. http://dx.doi.org/10.1037/0012-1649.43.1.186

Kushnir, T., Xu, F., & Wellman, H. M. (2010). Young children use statistical sampling to infer the preferences of other people. *Psychological Science*, *21*(8), 1134-1140. https://doi.org/10.1177/0956797610376652

Lawson, G. M., Hook, C. J., & Farah, M. J. (2018). A meta-analysis of the relationship between socioeconomic status and executive function performance among children. *Developmental Science*, *21*(2), e12529. https://doi.org/10.1111/desc.12529

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

Legare, C. H., Schult, C. A., Impola, M., & Souza, A. L. (2016). Young children revise explanations in response to new evidence. *Cognitive Development*, *39*, 45-56. https://doi.org/10.1016/j.cogdev.2016.03.003

Lucas, C. G., Bridgers, S., Griffiths, T. L., & Gopnik, A. (2014). When children are better (or at least more open-minded) learners than adults: Developmental differences in learning the forms of causal relationships. *Cognition*, *131*, 284-299. https://doi.org/10.1016/j.cognition.2013.12.010

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, *101*, 53-79. http://dx.doi.org/10.1037/0033-295X.101.1.53

Notaro, P. C., Gelman, S. A., & Zimmerman, M. A. (2001). Children's understanding of psychogenic bodily reactions. *Child Development, 72*, 444-459. https://doi.org/10.1111/1467-8624.00289

Perfors, A., Tenenbaum, J. B., Griffiths, T. L., & Xu, F. (2011). A tutorial introduction to Bayesian models of cognitive development. *Cognition*, *120*, 302-321. https://doi.org/10.1016/j.cognition.2010.11.015

Plebanek, D. J., & Sloutsky, V. M. (2017). Costs of selective attention: When children notice what adults miss. *Psychological Science*, *28*(6), 723-732. https://doi.org/10.1177/0956797617693005

Rhodes, M., & Wellman, H. (2013). Constructing a new theory from old ideas and new evidence. *Cognitive Science*, *37*(3), 592-604. https://doi.org/10.1111/cogs.12031

Rich, A. S., & Gureckis, T. M. (2018). The limits of learning: Exploration, generalization, and the development of learning traps. *Journal of Experimental Psychology: General, 147*(11), 1553–1570. https://doi.org/10.1037/xge0000466

Reverberi, C., Toraldo, A., D'Agostini, S., & Skrap, M. (2005). Better without (lateral) frontal cortex? Insight problems solved by frontal patients. *Brain*, *128*(12), 2882-2890. https://doi.org/10.1093/brain/awh577

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

Schapiro, A., & Turk-Browne, N. (2015). Statistical learning. *Brain Mapping: An Encyclopedic Reference*, *3*, 501-506.

Schulz, L. E., Bonawitz, E. B., & Griffiths, T. L. (2007). Can being scared cause tummy aches? Naive theories, ambiguous evidence, and preschoolers' causal inferences. *Developmental Psychology*, *43*, 1124-1139. http://dx.doi.org/10.1037/0012-1649.43.5.1124

Schulz, L. E., Goodman, N. D., Tenenbaum, J. B., & Jenkins, A. C. (2008). Going beyond the evidence: Abstract laws and preschoolers' responses to anomalous data. *Cognition*, *109*(2), 211-223. https://doi.org/10.1016/j.cognition.2008.07.017

Seiver, E., Gopnik, A., & Goodman, N. D. (2013). Did she jump because she was the big sister or because the trampoline was safe? Causal inference and the development of social attribution. *Child Development*, *84*(2), 443-454. https://doi.org/10.1111/j.1467-8624.2012.01865.x

Sim, Z. L., & Xu, F. (2017). Learning higher-order generalizations through free play: Evidence from 2-and 3-year-old children. *Developmental Psychology*, *53*, 642-651. http://dx.doi.org/10.1037/dev0000278

Slaughter, V., & Gopnik, A. (1996). Conceptual coherence in the child's theory of mind: Training children to understand belief. *Child Development*, *67*, 2967-88. https://doi.org/10.1111/j.1467-8624.1996.tb01898.x

Stahl, A. E., & Feigenson, L. (2015). Observing the unexpected enhances infants' learning and exploration. *Science*, *348*, 91-94. https://doi.org/10.1126/science.aaa3799

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

Thagard, P. (1992). *Conceptual Revolutions*. Princeton, NJ: Princeton University Press.

Thompson-Schill, S. L., Ramscar, M., & Chrysikou, E. G. (2009). Cognition without control: When a little frontal lobe goes a long way. *Current Directions in Psychological Science*, *18*(5), 259-263. https://doi.org/10.1111/j.1467-8721.2009.01648.x

van Schijndel, T. J., Visser, I., van Bers, B. M., & Raijmakers, M. E. (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

Walker, C. M., & Gopnik, A. (2014). Toddlers infer higher-order relational principles in causal learning. *Psychological Science*, *25*(1), 161-169. https://doi.org/10.1177/0956797613502983

Walker, C. M., Rett, A., & Bonawitz, E. (2020). Design drives discovery in causal learning. *Psychological Science*, *31*(2), 129-138. https://doi.org/10.1177/0956797619898134

Weinberger, A. B., Green, A. E., & Chrysikou, E. G. (2017). Using transcranial direct current stimulation to enhance creative cognition: interactions between task, polarity, and stimulation site. *Frontiers in Human Neuroscience*, *11*, 246. https://doi.org/10.3389/fnhum.2017.00246

Wente, A. O., Kimura, K., Walker, C. M., Banerjee, N., Fernández Flecha, M., MacDonald, B., Lucas, C., & Gopnik, A. (2019). Causal learning across culture and socioeconomic status. *Child Development*, *90*(3), 859-875. https://doi.org/10.1111/cdev.12943

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

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, *114*(2), 245–272. https://doi.org/10.1037/0033-295X.114.2.245

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature Protocols*, *1*(1), 297-301. https://doi.org/10.1038/nprot.2006.46

Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development*, *11*, 37-63. https://doi.org/10.1016/S0885-2014(96)90027-1

Zelazo, P. D., Müller, U., Frye, D., Marcovitch, S., Argitis, G., Boseovski, J., Chiang, J. K., Hongwanishkul, D., Schuster, B. V., Sutherland, A., & Carlson, S. M. (2003). The development of executive function in early childhood. *Monographs of the Society for Research in Child Development, 68*(3), i-151. https://www.jstor.org/stable/1166202