

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Maybe She'll Say Yes: How Young Learners Acquire and Apply Knowledge about Inconsistent Causal Relationships from Different Domains

Permalink

<https://escholarship.org/uc/item/79j7p447>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 47(0)

Authors

Lapidow, Elizabeth

Denison, Stephanie

Publication Date

2025

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Maybe She'll Say Yes: How Young Learners Acquire and Apply Knowledge about Inconsistent Causal Relationships from Different Domains

Elizabeth Lapidow (elapidow@uwaterloo.ca)

Department of Psychology, University of Waterloo

Stephanie Denison (stephanie.denison@uwaterloo.ca)

Department of Psychology, University of Waterloo

Abstract

Children are adept at learning the principles and properties governing their environment. However, this environment is often highly *inconsistent*: causes do not always bring about their effects; people do not always act according to their preferences. Past research shows that young causal learners readily reason *from* probabilistic evidence, but little is known as to how they reason *about* that evidence. This study presented preschoolers (N=114) with the behavior of three different causes—one consistently effective, one consistently ineffective, and one inconsistent—from one of three domains (social, mechanical, biological) and asked children to predict the future behavior of each. Children's predictions not only captured the different degrees of inconsistency observed in the evidence but also reflected differences in prior knowledge and expectations about inconsistency between domains. These results offer a novel, more nuanced look into early causal cognition and often-overlooked complexities of causal learning and reasoning in the real world.

Keywords: causal reasoning; cognitive development; probability; determinism; prediction; domain differences

Introduction

The world around us is often changeable and unreliable. While the goal of causal learning is to arrive at explanations or hypotheses that generalize from past to present situations, exceptions are inevitable (e.g., dropping plates usually, but not always, causes breaking, and waving at total strangers might, but usually won't, cause waving back). However, even young learners take these variations in stride when making causal inferences. Preschoolers readily infer others' preferences from inconsistent choices (e.g., choosing A over B three times and B over A once (Hu et al., 2015) and causes from probabilistic evidence (e.g., a cause brings about its effect 2/3 or 1/3 of the time, (Kushnir & Gopnik, 2005).

However, past work has not examined how learners represent or reason *about* these variations. What additional inferences (if any) do we make about causal relationships that give rise to probabilistic evidence? As adults, we seem to *have* additional intuitions or expectations about causal inconsistency. We would, for example, feel surprised if a friend responded to our greeting in a way they never had before. But that is nothing to what our surprise would be if a machine responded to our input in a way it never had before. While this difference feels intuitive, it draws on extensive experience with people and machines. How and when do these expectations form? Furthermore, expectations are far

more complex and nuanced than just broad differences between domains. We may believe that people are more variable than machines, but also that certain people—whether due to familiarity, temperament, or both—are more consistent than others. Similarly, whether we are more surprised by our go-to pain reliever or go-to coffee order not working as usual is likely a matter of complex individual differences in experience, mechanism knowledge, and more. Real-world causal learners likely have rich and complex beliefs about inconsistent causes, but to date, the causal learning literature does not provide an account of them. How prominent is their role in learners' reasoning about causal evidence? Can, for example, evidence about a particular relationship override what we believe is typical for its domain? If so, how much evidence is needed to revise an expectation that a particular cause is or is not likely to consistently bring about its effect?

The current study offers an initial investigation into these questions by examining 3- to 5-year-olds' reasoning about causal relationships that hold with different degrees of inconsistency. We employ a storybook task design, based on research demonstrating preschoolers' causal inferences from statistical information (e.g., Schulz et al., 2007), in which a character repeatedly interacts with three causes in hopes of producing the same outcome. One cause consistently produces the outcome (*Always* option), one consistently fails (*Never* option), and one is inconsistent and produces the outcome half of the time (*Sometimes* option). Participants are then asked to make a prediction about what each cause will do the next time the character tries. Our first question is whether and how children's predictions for the inconsistent cause will differ from their predictions for the two consistent causes. Past research (described below) suggests preschool-aged children will be adept at making causal inferences from patterns of evidence, but not how these inferences might differ when inconsistent and consistent patterns are directly compared.

We also consider the novel question of whether and how *domain* influences learners' reasoning about probabilistic evidence. Three versions of the storybook were used, between subjects, each with the same patterns of evidence, but differing in the type of causes involved: In the *Social* condition, the story described a character inviting three people to go play. In the *Mechanical* condition, he winds up three toys to make them go. And in the *Biological* condition, he puts out three different types of food for animals to eat.

This design allows us to investigate whether young learners have systematic expectations about inconsistency that are domain specific, as well as potential complex interactions between this prior knowledge and immediate evidence when making causal inferences.

Below, we review the literature on children's causal learning and reasoning from probabilistic events, as well as evidence suggesting potential domain differences in causal expectations and inference.

Probabilistic Causal Reasoning in Childhood

The basic fact of probabilistic causal learning in childhood is well established. Theories of cognitive development propose that knowledge takes the form of probabilistic causal models (e.g., Gopnik et al., 2004; Griffiths & Tenenbaum, 2009), and there is considerable evidence of children's causal learning from probabilistic evidence. However, while this literature is by no means small, there are contradictions about how readily young learners grasp evidence of inconsistent causes that hint at greater, unexplored complexities.

In an early investigation of children's causal learning, Kushnir and Gopnik (2005) examined preschoolers' inferences from probabilistic evidence. They introduced children to a 'blicket detector' box that played music when certain blocks are placed on top. Children watched the experimenter place two blocks individually on top of the detector multiple times, and were then asked to 'pick the best one to make it go.' The pairs of blocks differed in how often they activated the detector (e.g., Block A activated 3/3 times, Block B activated 1/3). Children reliably selected the more effective blocks, even when neither cause behaved consistently (e.g., 2/3 vs. 1/3). The apparent ease with which 4-year-olds in this task reasoned about probabilistic causes has been replicated in a number of other similar studies (e.g., Griffiths et al., 2011; Kushnir & Gopnik, 2007).

However, conclusions about this aspect of children's causal reasoning are mixed. Studies also show that 4-year-olds do not make causal generalizations about categories of objects from probabilistic data (Sobel et al., 2009) and struggle to reason from probabilistic evidence that suggests unlikely causal relationships (e.g., Kushnir & Gopnik, 2007), needs to be integrated with other information (Gualtieri et al., 2022), or when the difference in behavior between consistent and inconsistent causes is small (Sobel et al., 2009). While successful causal reasoning from probabilistic evidence is more consistently observed in 5- and 6-year-olds (e.g., Bonawitz et al., 2014; Sobel et al., 2009), one study finds that 3- to 5-year-olds resemble 'causal determinists,' explaining away stochastic evidence, rather than inferring the existence of an inherently inconsistent causal relationship (Schulz & Sommerville, 2006).

Findings about causal inference from inconsistent evidence prior to age 4 are even more puzzling. Waismeyer and colleagues (2015) find evidence to suggest even 2-year-olds can successfully infer the stronger of two causes from probabilistic evidence (2/3 vs. 1/3). However, the study employed highly unrealistic inconsistencies. Experimenters

showed children six demonstrations of each object (e.g., the 1/3 object was tried on the machine three times, followed by the 2/3 object, then each object three more times). Critically, the order in which each object succeeded and failed to activate the machine was *identical* both times it was observed. Thus, rather than creating an impression of stochasticity, the objects appeared to activate the machine in a consistent, repeating pattern. This raises the question of whether children's task performance is representative of their real-world causal learning and inference, which would typically not involve highly patterned structure.

While many factors likely contribute to these mixed findings (differences in study methodology, etc.), it also suggests underlying and complexities in how learners reason about inconsistent causes that have yet to be examined. In one particularly striking example (Sobel et al., 2009), 4-year-olds watched an experimenter try three objects from each of two different sets (e.g., square blocks and triangle blocks) on a detector. Then, the experimenter asked children which of two new objects (one from each set) would activate the detector. At first glance, children's performance on this task appears limited: 4-year-olds inferred which object would activate the machine from inconsistent success versus consistent failure (e.g., after observing 1 of 3 squares and 0 of 3 triangles activate the machine), but chose at chance when presented with inconsistent versus consistent success (e.g., after observing 2 of 3 squares and 3 of 3 triangles activate the machine).

However, this may not be due to children failing to track or make inferences from probabilistic causal evidence. Considering that unreliable causes are frequent in our real-world causal experience—experience which learners carry with them into experimental tasks (see Lapidow & Walker, 2022, 2024)—children's lack of preference between 3/3 and 2/3 causes suggests several other intriguing possibilities. Perhaps learners have a prior expectation that neither set of objects will activate the machine perfectly reliably, which three successful activation attempts are not enough to overturn. If so, learners might reasonably and rationally conclude that objects that 'nearly always work' are not meaningfully worse than objects that (at least so far) 'always work.' Alternatively, might the fact that only one block out of the total six doesn't activate the machine suggest that that failure was due to some external source (glitch in the detector, variation in where the experimenter placed the block, etc.) rather than differences in the inherent causal properties of the two block types? It is rich and complex questions such as these that causal learning research has yet to explore, and which the current investigation aims to undertake.

Expectations for Different Domains

The potential for inconsistency seems to be deeply embedded into our causal knowledge and reasoning: we can, for instance, believe that "dropping plates causes them to break" is true, without believing that every dropped plate will break (see Blanchard et al., 2018; Woodward, 1997 for full

discussion). Despite this, we also have a strong preference for more consistent causal relationships (Vasilyeva et al., 2018), and possibly even a tacit assumption that they should be consistent across contexts (e.g., Cheng & Lu, 2017). Similarly, we often, at least in explicit reasoning, seem to expect that causal relationships will be consistent. For adults, this is particularly true of mechanical domains. Such an expectation was at play in the opening example; a machine responding to an old input in a new way would be much more surprising than a person doing so. Similarly, we are often quick to jump to the conclusion that mechanical causes in particular are ‘broken’ when they behave inconsistently (e.g., if my go-to painkiller doesn’t relieve my headache, I am much less likely to immediately discard the bottle and buy a different brand than I am to start googling the cost of replacing a washing machine after mine fails to start).

It is notable, therefore, that the studies reviewed above present inconsistent causal evidence to children using novel *mechanical* systems. While this method has been invaluable for causal and cognitive development research, it provides a very particular, possibly limited, perspective on the complexities of real-world causal reasoning. One goal of the current study, therefore, is to examine how children reason about causal inconsistency when it is presented in different real-world contexts. There is some existing evidence to suggest that the role of domain in reasoning from and about inconsistencies is worth examining. In particular, research on early social cognition suggests that children easily take inconsistent evidence in stride when reasoning about the causes of others’ behavior. Four-year-olds readily and accurately infer both preferences (Hu et al., 2015) and trait-level dispositions (Seiver et al., 2013) from behavior that is not entirely consistent with those conclusions. For example, children infer that an agent likes a toy they chose five out of six times over more than one they chose seven out of eleven times, even though neither toy was always chosen. In addition, while complicated by the unrealistic patterning of inconsistent evidence, 2-year-olds’ performance in Waismeyer et al. (2015) replicates when the outcome is a person handing over marbles, rather than a detector activating (Waismeyer & Meltzoff, 2017).

The apparent ease with which children infer causes of people’s actions from inconsistent evidence is striking compared to the mixed results from mechanical detector-style tasks. Might learners be more tolerant of inconsistency when making inferences about the causes of social as opposed to physical events? To our knowledge, no studies have directly compared children’s probabilistic causal reasoning across domains (but see Schulz & Gopnik, 2004 and Weisberg et al., 2020 for comparisons with consistent causal evidence). One goal of the current study, therefore, is to provide the first direct comparison of early causal inference from inconsistent evidence across different domains. We chose three domains: *Social*, *Mechanical*, and *Biological*. We hypothesized that children’s predictions in the *Social* and *Mechanical* conditions would differ in (1) how the same patterns of evidence were interpreted between the

two domains and/or (2) whether the different patterns of evidence led to different predictions within a single domain. For example, young causal reasoners may already have an expectation that machines behave consistently, but people do not. If so, we would expect them to predict continued success and failure from the *Always* and *Never* options, respectively, in the *Mechanical* condition, but not necessarily in the *Social* condition. At the same time, we might also expect learners to weigh the differences between observed patterns of evidence differently depending on the domain. They might, for example, be more definite about their predictions of future success for *Always* and failure for *Never* in the *Mechanical* than *Social* condition. While we do not have similar a priori predictions about the nature of learners’ expectations or inferences in the *Biological* domain, this condition provides an important comparison. By presenting inconsistent events that are neither the workings of inanimate machines nor the behavior of human beings, this condition can potentially reveal whether any differences between the *Social* and *Mechanical* conditions are due to particularities in how children reason about either domain.

Methods

The design and analysis plan for this study were preregistered: <https://aspredicted.org/96m2-kxjx.pdf>. Data and example materials available: https://osf.io/kgb95/?view_only=8d74b23d95104bacbcf62feb1a0caa35.

Participants

The target n is 270, with 90 children from each age group (3-, 4-, and 5-year-olds) in each condition. Currently, data has been collected from 114 children (mean: 51.34 ms, range: 35.21–72.0 ms), with 45 in the *Social* (mean: 52.95 ms, range: 36.95–72.0 ms), 37 in *Mechanical* (mean: 52.13 ms, range: 36.95–71.47 ms), and 32 in *Biological* (mean: 48.15 ms, range: 35.21–71.9 ms). This allows for analysis of the effects of evidence and domain, treating age as a continuous variable. Children were recruited from schools and daycares (both public and private) and community centers (e.g., museums, food markets, and libraries) local to the university. We tested an additional 28 children, who were excluded and replaced for exceeding a preregistered prompt limit ($n = 11$), inattention or attrition ($n = 7$), and experimenter error ($n = 8$).

Materials

The task used six different storybooks, two for each condition (see OSF link for examples). Each book described the character interacting with three options, and the outcome for each option, on six different days. To ensure children were not simply using the first or last outcome observed to make their predictions, the two books for each condition showed a different sequence of behavior for the *Sometimes* option: either success on days 1, 4, and 6, or success on days 2, 3, and 5. The options (*Always*, *Sometimes*, *Never*) were paired to story elements described as ‘big,’ ‘middle/medium,’ and ‘small’ (e.g., “his little sister” or “the little car”). The pairing

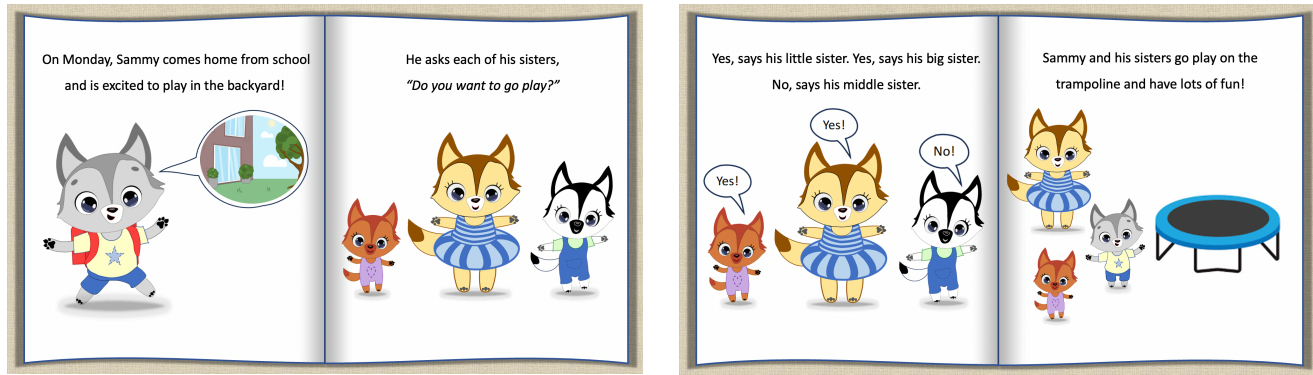


Figure 1: Example of storybook pages for a single day in the *Social* condition.

between options and story elements (e.g., which size car always goes) was counterbalanced. We also counterbalanced the order in which options were presented on a single page (e.g., which of the sisters was shown/named first) such that each option appeared twice in each position on two of the six days (the final day reused the order from the first day).

Procedure

Testing sessions began with a short warm-up, designed to introduce children to the style of the tasks and allow them to practice responding. The experimenter showed a picture of the storybook's main character, saying, "Today, we're going to learn all about my friend, Sammy." Children then had to respond to two simple forced-choice questions about the character (e.g., "Sammy's mom says he can either have a cookie or a cracker as a snack, which will Sammy choose?") and were corrected if they failed to give a valid response (e.g., no response, shrugging, trying to choose both options, and so on). The experimenter then brought out one of the six storybooks and read through it with the child. Each book followed the sequence shown in Figure 1 for a total of six days. If needed, the experimenters would pause reading the book to recall children's attention by pointing to one of the illustrations and asking, "What's that called?"

On the first six days of the storybook, the experimenter would read the page showing the character interacting with the three options, then immediately turn the page to reveal the outcomes. On the seventh day, the experimenter stopped after reading the interaction page and asked children to make a prediction about each option. In order to capture spontaneous nuance in children's responses (see below), the prediction question had two parts: an initial 'Yes/No' question, followed by a 'Definitely/Maybe' question, tailored to the initial response (e.g., One initial question in *Mechanical* was, 'Will the little car go, yes or no?' followed by, 'Will it definitely go/not go, or just maybe go/not go?'). If children did not immediately respond to either question, then the experimenter would repeat it (e.g., 'It's okay if you're not sure! Will the little sister definitely say yes or just maybe say yes?') followed by, if still no response, just the options ('Definitely or maybe?'). Children who gave non-responses after two prompts were excluded and replaced ($n = 11$).

Children's responses to the two test questions were combined and coded on an ordinal scale from 1 (*No; Definitely*), 2 (*No; Maybe*), 3 (*Yes; Maybe*), to 4 (*Yes; Definitely*). Some children, typically below 3 and a half, did not select an option on the 'definitely/maybe' question, but instead repeated their answer to the 'yes/no' question, often with added emphasis. If this happened, the experimenter would prompt once by asking, 'Definitely or maybe?'. If the child again repeated their initial answer, this was treated as equivalent to a 'definitely' response. At the same time, children across the age range often spontaneously included the word "maybe" in their responses, even before hearing a 'definitely/maybe' question (e.g., Many children in the *Social* condition said something like, "Maybe she'll say yes," in response to the 'yes/no' question for the first sister). This was treated as a 'maybe' response, and the experimenter would skip the 'definitely/maybe' question for this option.

Results

Figure 2 shows children's prediction scores for each option across the three conditions.

Following our preregistered analysis plan, we conducted an ordinal mixed effects logistic regression with prediction score (ordinal, between 1 and 4) as the dependent variable and predictor variables of Age (continuous), Condition, and Option, as well as the interaction between Option and Condition. Participant was also included as a random effect. The results of this model revealed no main effect of Age ($p = 0.55$), and significant, negative main effects for Option, with *Sometimes* (OR = 0.26 [95% CI: 0.11–0.62], $p < 0.01$) and *Never* (OR = 0.07 [95% CI: 0.03–0.18], $p < 0.001$) both more likely to receive lower prediction scores than *Always*. There was no significant main effect of Condition, but there was a marginal negative effect of *Social* compared to *Mechanical* (OR = 0.49 [95% CI: 0.22–1.08], $p = 0.08$), suggesting that at least some options in the *Mechanical* condition received higher ratings than those in the *Social* condition. In addition, there were significant interactions between Option and Condition, suggesting that the effect of past evidence on children's predictions varied depending on the domain of that evidence. Specifically, there were significant positive effects of the interaction between *Never* and *Social* (OR = 4.88 [95%

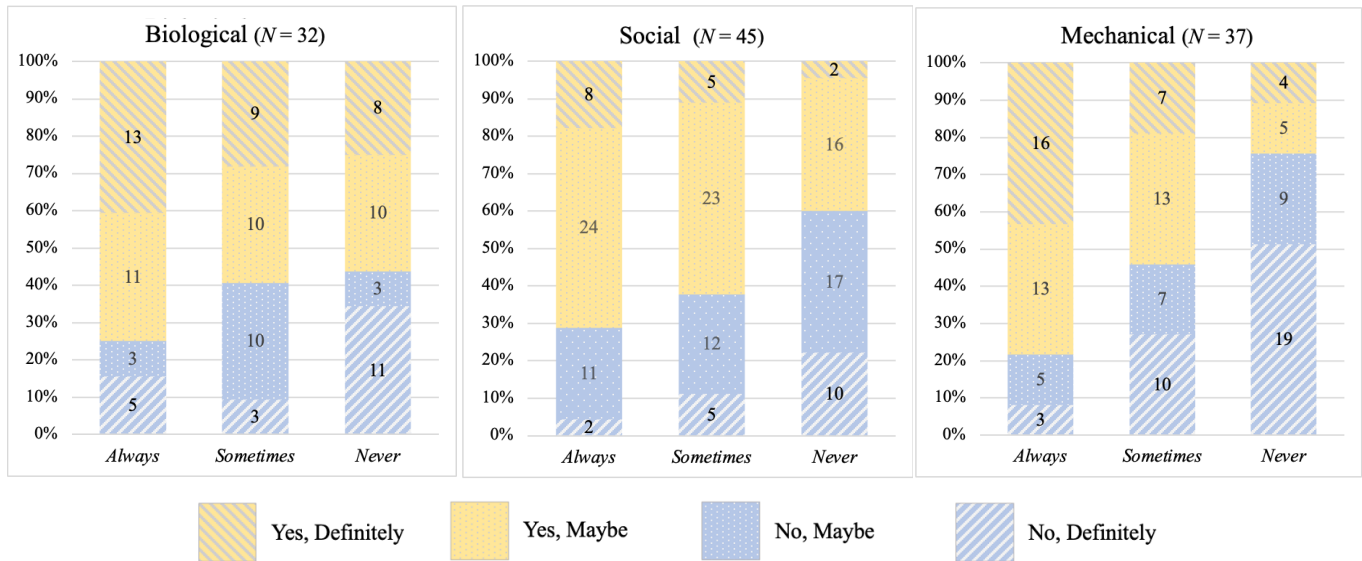


Figure 2: Number of children giving each prediction across each of the three conditions.

CI: 1.57–15.14], $p = 0.01$) and between *Never* and *Biological* (OR = 4.58 [95% CI: 1.24–16.91], $p = 0.02$), as well as a marginal positive effect for *Sometimes* and *Social* (OR = 2.68 [95% CI: 0.87–8.24], $p = 0.08$). These results suggest that, while *Sometimes* and *Never* received lower ratings compared to *Always*, the ratings for these options were often higher in *Social* and *Biological* than *Mechanical* conditions—which would be consistent with a greater number of “Definitely No” responses when reasoning about machines, than about animate beings like people and animals.

Recall that our hypothesis was that young children would combine inferences from patterns of causal evidence with different prior expectations of inconsistency in different domains. Therefore, we followed the regression analysis with chi-square tests of whether prediction scores differed between options, both across and within conditions. First, children’s responses showed a domain-general sensitivity to the evidence observed: Prediction scores for *Always*, *Sometimes*, and *Never* differed significantly from one another in the *Biological*, $X^2(6, N = 96) = 13.06, p = 0.04$, and *Mechanical*, $X^2(6, N = 111) = 26, p < 0.001$, conditions. In the *Social* condition, the difference between options was marginally significant, $X^2(6, N = 135) = 12.72, p = 0.05$. Second, there was also evidence of domain differences in children’s reasoning. Despite observing identical patterns of evidence, predictions for *Never* differed significantly across the three conditions (*fisher’s exact test*, $p < 0.01$) while differences in predictions for the *Always* (*fisher’s exact test*, $p = 0.05$) and *Sometimes* (*fisher’s exact test*, $p = 0.12$) options did not reach significance.

Discussion

The primary goal of the current investigation is to serve as a first step towards a clearer and more complete picture of early reasoning about inconsistent causes. Our results suggest that young learners are sensitive to patterns of inconsistent versus

consistent causal evidence *and* have rich prior expectations about inconsistency in different domains—both of which combine in how they reason from and about the behavior of inconsistent causal relationships (Figure 2).

Children’s predictions showed clear sensitivity to the different behaviors they observed from the three options. Across conditions, children made “Definitely No” predictions most often for *Never* and “Definitely Yes” predictions most often for *Always*. Indeed, there were statistically significant differences between the predictions for the three options in both *Mechanical* and *Biological* conditions. This difference was marginal in the *Social* condition, which is likely due to influences of domain-specific expectations on reasoning rather than any failure to attend to the evidence. Instead, the results suggest that, compared to machines and animals, differences between the past behaviors of people does not lead to similarly firm differences in learners’ predictions for their future behavior: “Maybe” predictions made up 76% of children’s responses in the *Social* condition, compared to 49% in *Biological* and 47% in *Mechanical*. At the same time, however, children did not ignore the evidence before them and reason purely from prior expectations: “Maybe” predictions were the most common response for inconsistent causal relationships (i.e., for the *Sometimes* option), regardless of the domain of that relationship.

Children’s success at tracking and making appropriate inferences about differently consistent causes contrasts with past findings that young children, particularly under the age of 5, struggle with patterns of inconsistent causal evidence (e.g., Sobel et al., 2009; Schulz & Somerville, 2006). However, the fact that even 3-year-olds demonstrated sophisticated and sensitive probabilistic causal reasoning in our task does not invalidate these past findings, but rather opens numerous avenues for future investigation. For example, it is possible that asking children for graded

predictions, rather than binary judgements, allowed for a more sensitive assessment of their understanding of inconsistent causes. Consider, for example, children's apparent failure to identify the stronger cause when 2/3 was pitted against 3/3 in Sobel et al. (2019). If the current task had asked children to choose whether *Sometimes* or *Always* was the better option for the character to try next, rather than to make predictions about them, performance may have looked similar to the earlier study. In the same way, children predominately giving "maybe" responses for the *Sometimes* option, particularly in the *Mechanical* condition, seems at odds with the suggestion that young learners refuse to accept inherently stochastic causes (Schulz & Sommerville, 2006). If, however, it were possible to probe the *reasons* children have in mind for their predictions, we might find similar appeals to external variables to explain why some causes appear inconsistent (e.g., that the character just didn't wind two of the cars up enough on some days). Other studies currently underway in our lab examine the role internal and external sources of variability might play in learners' reasoning about causal inconsistency.

This study is also the first to directly compare children's ability to reason about the same patterns of inconsistent causal evidence across different domains. Our results provide substantial evidence for an influence of domain-specific prior knowledge and expectations: Children did not reason about *Always*, *Sometimes*, and *Never* causes in the same way across the *Social*, *Mechanical*, and *Biological* stories. The vast majority of children predicted that a mechanical toy that had never previously worked would 'definitely not' work when next tried. In contrast, predictions about the future behavior of a sister who had consistently refused to come play and a birdseed that had consistently failed to attract birds were far less definite. Similarly, a roughly equal number of children predicted that a toy car or birdseed that had always previously worked would 'definitely' or 'maybe' work again, but far and away the most common prediction for the sister who always agreed to go play was that she'd only 'maybe' do so when next asked. Indeed, 'maybe' was the predominant prediction for all three options in the *Social* condition, suggesting children consider the future behavior of other agents more potentially variable or uncertain than causes from other domains that have behaved with the same degree of (in)consistency.

The apparent influence of prior expectations on reasoning about causal evidence is particularly intriguing for several reasons. First, it offers a potential explanation for why inconsistent evidence seems less of a stumbling block to young learners' reasoning in the social domain (e.g., Hu et al., 2015; Seiver et al., 2013). Furthermore, the difference between social and nonsocial causal relationships observed here is intriguingly consistent with other findings in psychology and philosophy. For instance, adult reasoners expect events in the social world to have more causes and more complex causal structures than those in the physical world (e.g., we think it more likely that a person bursting into tears is the result of numerous contributing factors

converging together, whereas a tree falling is more likely the result of a linear sequence of single events, Strickland et al., 2017). In addition, philosophies of free-will and determinism have long included the concept of *compatibilism*, the notion that belief in a deterministic universe is compatible with belief in the free-will (and thus, responsibility) in the actions of agents (see McKenna & Coates, 2024). There is also some evidence to suggest that lay intuitions are compatibilist (e.g., Nichols & Knobe, 2007). It is beyond the scope of the current study to say whether children's predictions reflected a belief that human behavior is fundamentally variable in a way the physical world is not, or resulted from our natural access to the complexity of factors involved in why people do things. However, both possibilities would be consistent with what the present results reveal about causal reasoning. Namely, that our perceptions and/or intuitions about the causes of human behavior are different—and often involve greater attributions of possibility and unpredictability—than our reasoning about causal relationships in other domains.

One limitation of the current design is that we did not assess children's predictions *prior* to observing the patterns of evidence. Doing so would have allowed us to capture the baseline strength of their beliefs about the likelihood of inconsistency in these different domains. This choice was intentional, as requesting a prediction for each option when there are no apparent differences between them and eliciting repeated responses from children before and after evidence might have introduced other confounds. This means we cannot directly estimate how learners integrate and trade off between current evidence and prior expectations within the task. Studies are currently planned to examine not only this question, but also to compare how this complex reasoning about the interplay of expectation and evidence might develop and differ between children and adults.

The current study offers significant novel insights for research in both cognitive development and causal cognition. It highlights the role of complex conceptual knowledge in how learners understand and interpret probabilistic causal evidence, even when the strength or stability observed in that evidence is identical. Our hope is that this investigation can serve as a first step towards a larger examination into these and other critical and often-overlooked complexities inherent to real-world causal learning and reasoning.

Acknowledgments

The authors would like to thank the invaluable undergraduate research assistants at the Developmental Learning Lab for their work on data collection and the members of the University of Waterloo Psychology Department for their feedback. This research was supported by funding from the Social Sciences and Humanities Research Council of Canada and the Jacobs Foundation Young Scholar Project.

References

- Blanchard, T., Vasilyeva, N., & Lombrozo, T. (2018). Stability, breadth and guidance. *Philosophical Studies*, 175(9), 2263–2283.

- Bonawitz, E., Denison, S., Gopnik, A., & Griffiths, T. L. (2014). Win-Stay, Lose-Sample: A simple sequential algorithm for approximating Bayesian inference. *Cognitive Psychology*, 74, 35–65.
- Cheng, P. W., & Lu, H. (2017). Causal invariance as an essential constraint for creating representation of the world: Generalizing the invariance of causal power. *The Oxford handbook of causal reasoning*, 65–84.
- Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A Theory of Causal Learning in Children: Causal Maps and Bayes Nets. *Psychological Review*, 111(1), 3–32.
- Griffiths, T. L., Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2011). Bayes and blickets: Effects of knowledge on causal induction in children and adults. *Cognitive Science*, 35(8), 1407–1455.
- Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological Review*, 116(4), 661–716.
- Gualtieri, S., Attisano, E., & Denison, S. (2022). Young children’s use of probabilistic reliability and base-rates in decision-making. *PLOS ONE*, 17(5), e0268790.
- Hu, J., Lucas, C. G., Griffiths, T. L., & Xu, F. (2015). Preschoolers’ understanding of graded preferences. *Cognitive Development*, 36, 93–102.
- Kushnir, T., & Gopnik, A. (2005). Young children infer causal strength from probabilities and interventions. *Psychological Science*.
- 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(1), 186.
- Lapidow, E., & Walker, C. M. (2022). Rethinking the “gap”: Self-directed learning in cognitive development and scientific reasoning. *WIREs Cognitive Science*, 13(2), e1580.
- Lapidow, E., & Walker, C. M. (2024). Learners’ causal intuitions explain behavior in control of variables tasks. *Developmental Psychology*, 60(12), 2258–2273.
- McKenna, M., & Coates, D. J. (2024). Compatibilism. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Summer 2024). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/sum2024/entries/compatibilism/>
- Nichols, S., & Knobe, J. (2007). Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions. *Noûs*, 41(4), 663–685.
- Schulz, L. E., Bonawitz, E. B., & Griffiths, T. L. (2007). Can being scared cause tummy aches? Naïve theories, ambiguous evidence, and preschoolers’ causal inferences. *Developmental Psychology*, 43(5), 1124–1139.
- Schulz, L. E., & Gopnik, A. (2004). Causal learning across domains. *Developmental Psychology*, 40(2), 162.
- Schulz, L. E., & Sommerville, J. (2006). God Does Not Play Dice: Causal Determinism and Preschoolers’ Causal Inferences. *Child Development*, 77(2), 427–442.
- 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.
- Sobel, D. M., Sommerville, J. A., Travers, L. V., Blumenthal, E. J., & Stoddard, E. (2009). The Role of Probability and Intentionality in Preschoolers’ Causal Generalizations. *Journal of Cognition and Development*, 10(4), 262–284.
- Strickland, B., Silver, I., & Keil, F. C. (2017). The texture of causal construals: Domain-specific biases shape causal inferences from discourse. *Memory and Cognition*.
- Vasilyeva, N., Blanchard, T., & Lombrozo, T. (2018). Stable causal relationships are better causal relationships. *Cognitive Science*, 42(4), 1265–1296.
- Waismeyer, A., & Meltzoff, A. N. (2017). Learning to make things happen: Infants’ observational learning of social and physical causal events. *Journal of Experimental Child Psychology*, 162, 58–71.
- Waismeyer, A., Meltzoff, A. N., & Gopnik, A. (2015). Causal learning from probabilistic events in 24-month-olds: An action measure. *Developmental Science*, 18(1), 175–182.
- Weisberg, D. S., Choi, E., & Sobel, D. M. (2020). Of Blickets, Butterflies, and Baby Dinosaurs: Children’s Diagnostic Reasoning Across Domains. *Frontiers in Psychology*, 11, 2210.
- Woodward, J. (1997). Explanation, invariance, and intervention. *Philosophy of Science*, 64, S26–S41.