Title
Expectations of Causal Determinism in Causal Learning

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 44(44)

Authors
Dinh, Phuong (Phoebe) Ngoc
Danks, David

Publication Date
2022

Peer reviewed

Expectations of Causal Determinism in Causal Learning

Phuong (Phoebe) Ngoc Dinh (pdinh@andrew.cmu.edu)
Department of Psychology, 5000 Forbes Avenue
Pittsburgh, PA 15213 USA

David Danks (ddanks@ucsd.edu)
Department of Philosophy and The Halıcıoğlu Data Science Institute, 9500 Gilman Drive
La Jolla, CA 92093 USA

Abstract

Causal learning is shaped by people's prior beliefs, including their expectations. In this paper, we specifically examine expectations of determinism: do they vary with perceptual features of physical causal events, and how do they influence subsequent causal learning from data? We show that perceptual features lead adults to different expectations of determinism for different causes of launching (Exps. 1A & 1B). Those expectations lead to significant differences in responses to causal "failures"; that is, we show a difference in violation-of-expectation effect after a failed launch (Exp. 2). Actual data can reduce or eliminate the impact of these expectations, but they do not override the effect of perceptual features (Exp. 3). Overall, spatiotemporal contiguity cues and expectation of determinism have similar effects on causal learning outcomes, but neither is fully reducible to the other.

Keywords: causal learning; determinism; sequential learning paradigm; violation of expectation

Determinism and Causal Learning

Causal learning—the ability to identify and represent causal relations given data—is crucial to human cognition. Many theories of causal learning focus on nondeterministic settings and largely predict learning to be gradual (e.g., Bonawitz et al., 2014; Bramley et al., 2017; Cheng, 1997; Fernbach & Sloman, 2009). Yet causal learning need not be incremental: if people expect a causal relation to be deterministic, then they can learn from just one case (Michotte, 1963; Scholl & Nakayama, 2002). Determinism—the general idea that an effect is entirely and reliably determined by one or a few causes—is a potent cue for causal learning (Deverett & Kemp, 2012; Lu et al., 2008). As such, people often form expectations about how deterministic a causal relation is (Frosch & Johnson-Laird, 2011; Yin & Sun, 2021). Notably, four-year-olds resist the idea that causes can be inherently stochastic (Schulz & Sommerville, 2006).

In this paper, we examine three different questions about interactions in causal learning between expectations of determinism, perceptual features, and sequences of data:

1. <u>Exps. 1A & 1B</u>: Do expectations of determinism vary based on perceptual features (within a single domain)?
2. <u>Exp. 2</u>: Do differences in expectation of determinism lead to different causal learning from nondeterministic sequences of data?
3. <u>Exp. 3</u>: How do manipulations of expectations of determinism affect later sequential causal learning?

Experiments 2 and 3 use a sequential causal learning paradigm in which participants see a series of cases and repeatedly evaluate the power of the apparent causes. This design enables us to study the causal learning trajectory. In contrast with most prior sequential causal learning studies (e.g., Danks & Schwartz, 2006; Marsh & Ahn, 2009), we focus on a domain—physics—where deterministic causal relations are often expected (Yeung & Griffiths, 2015; Yin & Sun, 2021), though not necessary. This focus enables us to manipulate determinism expectations and use a violation-of-expectation paradigm, thereby expanding our understanding of the role of determinism in sequential causal learning.

Experiments 1A & 1B

Many studies on determinism in causal learning has varied expectations by using situations from different domains, thereby confounding (expectations of) determinism and domain effects (e.g., Strickland et al., 2017; Yeung & Griffiths, 2015). Our first experiment examines the role of perceptual features on people's expectations of determinism in the single domain of physical causation.

Method

Participants Two independent samples of 63 adults each were recruited for Experiments 1A and 1B on Amazon Mechanical Turk (MTurk). All participants had a HIT approval rating of >90% and were paid $2.50.

Materials & Design All participants watched six animated launching events (hosted on Youtube) in random order, each lasting 4-5 s with 0.67 s blue screens at the start and end. (All videos can be found at https://osf.io/uqv5a/.) All events showed a ball starting to move, but in different ways:

- *launch*: A red ball is stationary in center. A blue ball enters from left and contacts the red ball. The red ball then moves rightward and the blue ball remains in center.
- *rebound*: Identical to *launch* except that the blue ball rolls back (left) slowly for 1.3 s as if it rebounded.
- *delay*: Identical to *launch* except that the red ball does not move until 1 s after the blue ball contacts it.
- *gap*: Identical to *launch* except the blue ball stops before contact with the red ball, which then begins moving.
- *gradual*: A yellow ball is stationary in center. Its color changes to green over 1.4 s, and then moves rightward.

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society*.

- *blink*: A peach-colored ball is stationary in center. Its color flickers rapidly six times (in 1.4 s) between peach and purple, and then stays purple and moves rightward.

Delay and gap were chosen because they disrupted temporal and spatial contiguity, respectively, between the cause and the effect. Gradual and blink removed the external agent yet preserved the spatial and temporal contiguity between the candidate cause and effect (Schlottmann & Shanks, 1992).

Experiments 1A and 1B differed based on the question asked after each event about the obvious candidate cause:

- Exp. 1A: Rate the degree to which [candidate cause] seemed like a convincing cause of the ball's motion (0 = *not at all*, 100 = *absolutely*, in increments of 10)
- Exp. 1B: Estimate the number of cases in which the stationary ball would move, given 100 cases in which the cause was present (0 to 100 scale in increments of 1)

Procedure Participants joined on MTurk and were directed to Qualtrics. After consent, they completed a check of the YouTube console and watched the six events, answering the two questions after each event. Participants completed an instructional manipulation check and read a debriefing form.

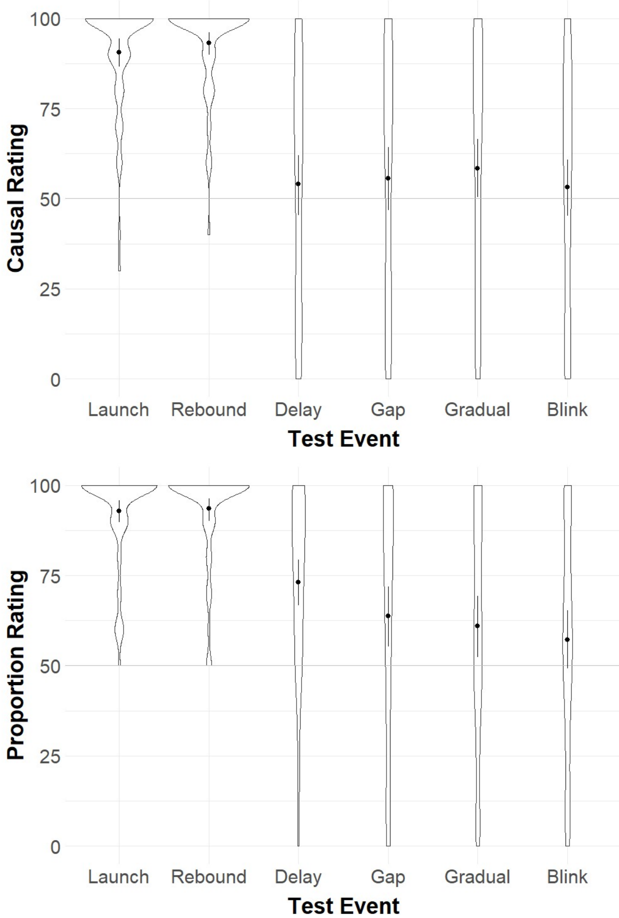

Figure 1: Ratings of different launch events. Error bars: 95% CI.

Results

Although different rating/estimation questions were used in the two versions of Experiment 1, they provide the same qualitative pattern of results (Figure 1). Linear regressions were run for each question-type with the type of cause as predictor (and *launch* as the reference category).

In Experiment 1A, there was no significant difference between causal perception ratings of *launch* and *rebound* (B_{rebound} = 2.54, p = 0.64, 95% CI = [-2.16, 7.43]). In contrast, participants gave significantly lower causal perception ratings to the apparent cause in the other four events: *delay* (B_{delay} = -36.5, p < .001, 95% CI = [-45.1, -27.3]); *gap* (B_{gap} = -35.1, p < .001, 95% CI = [-44.3, -25.7]); *gradual* (B_{gradual} = -32.2, p < .001, 95% CI = [-41.7, -22.8]) and *blink* (B_{blink} = -37.5, p < .001, 95% CI = [-46.6, -28.6]).

Experiment 1B had the same result pattern: no significant difference between proportion ratings of *launch* and *rebound* (B_{rebound} = 0.635, p = 0.54, 95% CI = [-4.08, 5.27]), but significantly lower proportion ratings for *delay* (B_{delay} = -19.7, p < 0.001, 95% CI = [-26.5, -12.6]); *gap* (B_{gap} = -29.0, p < 0.001, 95% CI = [-37.9, -20.3]); *gradual* (B_{gradual} = -31.9, p < 0.001, 95% CI = [-40.7, -23.0]); and *blink* (B_{blink} = -35.7, p < 0.001, 95% CI = [-44.7, -26.9]).

Discussion

Results, whether ratings of perceptual realism or a more traditional causal strength measure, echo prior research. A rebound effect did not significantly alter participants' judgments, but a delay or gap between the agent and its recipient lowered causal judgments (Michotte, 1963). Most importantly, a color change, whether gradual or sudden, was deemed a weak cause of motion of a stationary ball even when it was the only apparent cause in the event.

Experiment 2

In Experiments 1A and 1B, participants judged *launch* to be a highly effective, almost-deterministic cause, while *blink* appeared to be a less powerful and less deterministic cause (though not ineffectual). We thus focused on those two types of events in Experiments 2 and 3. In Experiment 2, we tested if participants' expectations about causal determinism affect causal learning over multiple trials, particularly depending on whether those expectations were supported or violated.

Method

Participants A sample of 98 adults was recruited on MTurk, all with a HIT approval rating >95%, and physical location in the U.S. Participants received $2.50 for their time. Fifteen participants were excluded for failing an attention check, and two for misinterpreting rating scales, resulting in N=81.

Materials & Design Participants watched two series, each with twelve animated (4 s) events of the same type. One series had *launch* events and the other had *blink* events.

If the series was *deterministic*, then the cause was always followed by the stationary ball moving. If the series was *probabilistic*, then the cause only led to the motion of the stationary ball on 75% (9/12) of the cases. In all probabilistic series, participants observed four instances of successful causation, followed by a "failure" case on trial 5. (The other failures were at trial 7, then either 8 or 11.) All participants saw one deterministic and one probabilistic series. Presentation order was counterbalanced across participants.

For each series, participants answered two questions after events 1, 3, 5, 7, 9, 11, and 12. Participants rated the extent to which the apparent cause (contact with the moving ball or color change) made the stationary ball move (-100 = *[cause] prevents movement*, 0 = *No relationship between [cause] and movement*, 100 = *[cause] causes movement* in increments of 10). They also estimated the number of cases in which they would expect the stationary ball to move, given 100 cases in which [cause] was present (0 to 100 scale in increments of 1). For both questions, participants were explicitly instructed to consider *all* of the events that they had seen in that series.

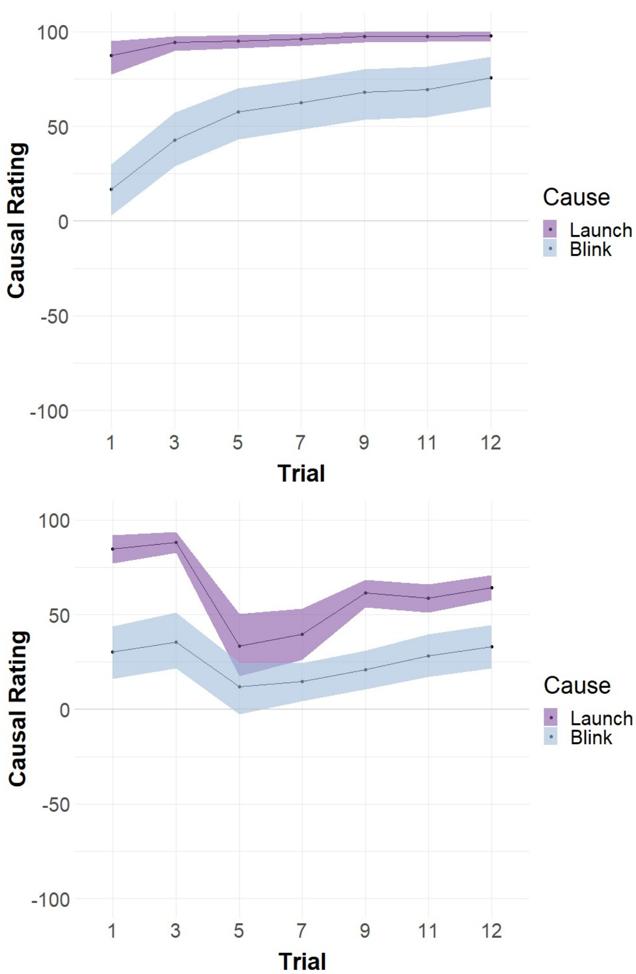

Figure 2: Mean causal ratings for deterministic (top) and probabilistic (bottom) sequences. Error ribbon: 95% CI.

Results

Linear mixed-effect regressions (LMEs) were used to compare both kinds of judgments between trials 1, 5 (first failure case in the probabilistic series), and 12 (final case). These comparisons provide a description of the pattern of data rather than a complete statistical model. In all models, trial number, cause-type, trial-cause interaction, and order were fixed factors, and participant ID was the random factor. Trial 1 and *blink* were the reference categories. Analyses of both types of judgments revealed similar patterns.

Causal Ratings Causal ratings for the deterministic series (Figure 2) were as expected. *launch* was higher than *blink* at trial 1 (B_{blink} = -82.6, SE = 10.4, $t(154)$ = -7.98, $p < .001$, 95% CI = [-102, -62.7], $R^2_{partial}$ = .374). Ratings increased across trials, both at trial 5 ($B_{5,\text{ blink}}$ = 33.9, SE = 10.4, $t(154)$ = 3.27, p = .0014, 95% CI = [13.9, 53.8], $R^2_{partial}$ = .118) and trial 12 ($B_{12,\text{ blink}}$ = 56.5, SE = 10.4, $t(154)$ = 5.44, $p < .001$, 95% CI = [36.5, 76.4], $R^2_{partial}$ = .219). This increase is largely driven by *blink*, though *launch* ratings are likely at ceiling.

Causal ratings for the probabilistic series revealed three main effects of trial numbers and type of cause. Relative to ratings at trial 1, participants' causal ratings for *launch* dropped significantly at trial 5 (B_5 = -58.0, SE = 10.7, $t(154)$ = -5.41, $p < .001$, 95% CI = [-78.6, -37.4], $R^2_{partial}$ = .217) and marginally at trial 12 (B_{12} = -19.0, SE = 10.7, $t(154)$ = -1.77, p = .0783, 95% CI = [-39.6, 1.60], $R^2_{partial}$ = .017). That is, participants' causal ratings for *launch* never recovered to their initial level. Participants also gave lower causal ratings for *blink* than *launch* at trial 1 (B_{blink} = -62.0, SE = 12.7, $t(198)$ = -4.87, $p < .001$, 95% CI = [-86.4, -37.6], $R^2_{partial}$ = 0.159).

Notably, the drop in causal rating from trial 1 to trial 5 in the probabilistic series was smaller for *blink* than *launch* ($B_{5,\text{blink}}$ = 40.5, SE = 15.2, $t(154)$ = 2.67, p = .00835, 95% CI = [11.4, 69.6], $R^2_{partial}$ = .059). This finding is consistent with a violation-of-expectation effect: participants expected determinism for *launch* (but not *blink*), and the first failure case violated that expectation. A similar, though smaller, interaction effect was found between trial 12 and type of cause: The drop in causal rating from trial 1 to 12 was larger for *launch* than *blink* ($B_{12,\text{blink}}$ = 26.5, SE = 15.2, $t(154)$ = 1.75, p = .0824, 95% CI = [-2.64, 55.6], $R^2_{partial}$ = .031). That is, the nondeterminism impacted *launch* judgments more than *blink*, even when the statistics of the event series were identical.

Proportion Ratings Due to limited space, we omit the graphs of proportion ratings as they are qualitatively the same as Figure 2; for completeness, the graphs are provided at https://osf.io/f3bjq/. In the deterministic series, proportion ratings were higher for *launch* than *blink* at trial 1 (B_{blink} = -44.7, SE = 7.00, $t(158)$ = -6.38, $p < .001$, 95% CI = [-58.1, -31.2], $R^2_{partial}$ = .285). Proportion ratings overall increased significantly from trial 1 to trial 12 (B_{12} = 10.4, SE = 4.99, $t(154)$ = 2.082, p = .0390, 95% CI = [0.798, 20.0], $R^2_{partial}$ = .284), primarily due to *blink* ($B_{\text{blink},12}$ = 22.2, SE = 7.14, $t(154)$ = 3.106, p = .00226, 95% CI = [8.45, 35.9], $R^2_{partial}$ = .107)

The probabilistic series showed three main effects. Proportion ratings dropped from trial 1 to trial 5 (B_5 = -20.9, SE = 5.85, $t(154)$ = -3.57, p = .000473, 95% CI = [-32.1, -9.65], R^2_{partial} = .122) and to trial 12 (B_{12} = -14.4, SE = 5.85,

$t(154) = -2.45$, $p = .0153$, 95% $CI = [-25.6, -3.10]$, $R^2_{\text{partial}} = .024$). Participants overall gave lower proportion ratings for *blink* ($B_{\text{blink}} = -34.7$, $SE = 8.00$, $t(160) = -4.33$, $p < .001$, 95% $CI = [-50.0, -19.3]$, $R^2_{\text{partial}} = .144$). Notably, the magnitude of drop in proportion ratings from trial 1 to trial 12 was modulated by the type of cause, as the difference was smaller for *blink* than *launch* ($B_{12,\text{blink}} = 17.1$, SE = 8.28, $t(154) = 2.067$, $p = .0405$, 95% CI = [1.19, 33.0], $R^2_{\text{partial}} = .083$).

Discussion

Experiment 2 showed that expectations of determinism influence causal learning trajectories, as those differed for series with the same statistics but different expectations. If a cause was expected to be deterministic, then causal and proportion ratings were near ceiling so long as there were no violations. A failure case, however, produced a sharp decline of both causal and proportion ratings. In contrast, if the cause was expected to be probabilistic, then causal and proportion ratings started relatively lower, increased gradually with each success, and decreased gradually with each failure.

Experiment 3

Experiment 2 could not determine if perceptual features and determinism expectations separately affect causal learning, or whether features produce expectations which influence learning. Experiment 3 used an "alien object" cover story to exogenously manipulate expectations of determinism, since participants would not necessarily assume that alien objects behave like those on Earth. By independently manipulating perceptual features and expectations of determinism, we tested if the effect of the former was solely through the latter.

Method

Participants We recruited 440 participants on MTurk with a HIT approval rating >95% and physical location in the U.S. Participants received $2.00. Data from 113 participants were removed for failing an attention check, resulting in N=327.

Materials & Design Before seeing any cases, participants read a brief story about alien objects being brought to Earth for study. No one has figured out what these objects are, though there are some preliminary reports. The objects do not normally move, but in response to certain internal or external changes, they behave either in "regular and consistent ways" (deterministic expectation condition), or in "surprising and inconsistent ways" (probabilistic expectation condition). A manipulation recall check was used to ensure that participants remembered the expectation; those who failed the check were reminded of it before seeing the sequence of cases.

We chose a between-participants design to reduce fatigue and interference. Each participant saw only one series of 12 events from one of 8 conditions: Expectation {deterministic, probabilistic} × Sequence {deterministic, probabilistic} × Cause {*launch*, *blink*}. In the probabilistic sequence, failure cases were trials 5, 7, and 11. The causal and proportion rating questions and scales from Experiment 2 were used.

Results

The basic data analysis mirrored Experiment 2 but included consistency (between expectation and sequence-type) in the LME models. However, the data analysis was complicated by the fact that participants who passed the manipulation recall check (N = 108, 93 for deterministic, probabilistic series) responded differently in several ways from those who failed the check and had to be reminded (N = 54, 72). We thus report separate analyses based on manipulation recall, and later summarize key differences between those two groups.

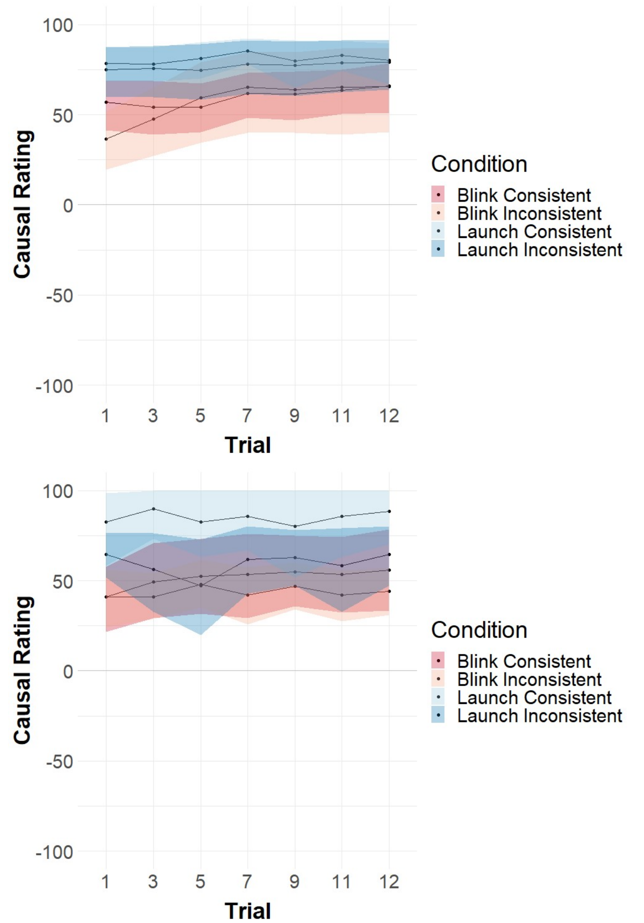

Figure 3: Mean causal ratings for deterministic sequence for participants who passed (top) and failed (bottom) the manipulation recall check. Error ribbon: 95% CI.

Causal Ratings For the deterministic series, participants who passed the manipulation recall check (Figure 3 top) gave lower causal ratings for *blink* than *launch*, including at trial 1 ($B_{\text{blink}} = -21.8$, $SE = 10.7$, $t(185) = -2.034$, $p = .0433$., 95% $CI = [-42.5, -1.09]$, $R^2_{partial} = .074$). *Blink* causal ratings increased from trial 1 to trial 5, and this increase was greater with inconsistent instruction indicating that the alien objects behaved unreliably ($B_{\text{blink,5,inconsistent}} = 28.2$, $SE = 14.2$, $t(208) = 1.99$, $p = .0479$, 95% $CI = [0.820, 55.5]$, $R^2_{partial} = .019$).

For that series, participants who failed the recall check (Figure 3 bottom) only gave significantly lower causal ratings for *blink* than *launch* at trial 1 (B_{blink} = -41.7, SE = 14.2, t(89) = -2.94, p = .0042, 95% CI = [-68.7, -14.7], $R^2_{partial}$ = .113).

For the probabilistic series, participants who passed the manipulation recall (Figure 4 top) gave *launch* causal ratings that dropped significantly from trial 1 to trial 5, the first failure (B_5 = -39.0, SE = 14.4, t(178) = -2.70, p = .0076, 95% CI = [-66.8, -11.2], $R^2_{partial}$ = .117). Causal ratings were lower for *blink* than *launch* at trial 1, though this main effect was only marginally significant (B_{blink} = -28.0, SE = 14.5, t(237) = -1.94, p = .054, 95% CI = [-55.8, -.191], $R^2_{partial}$ = .049).

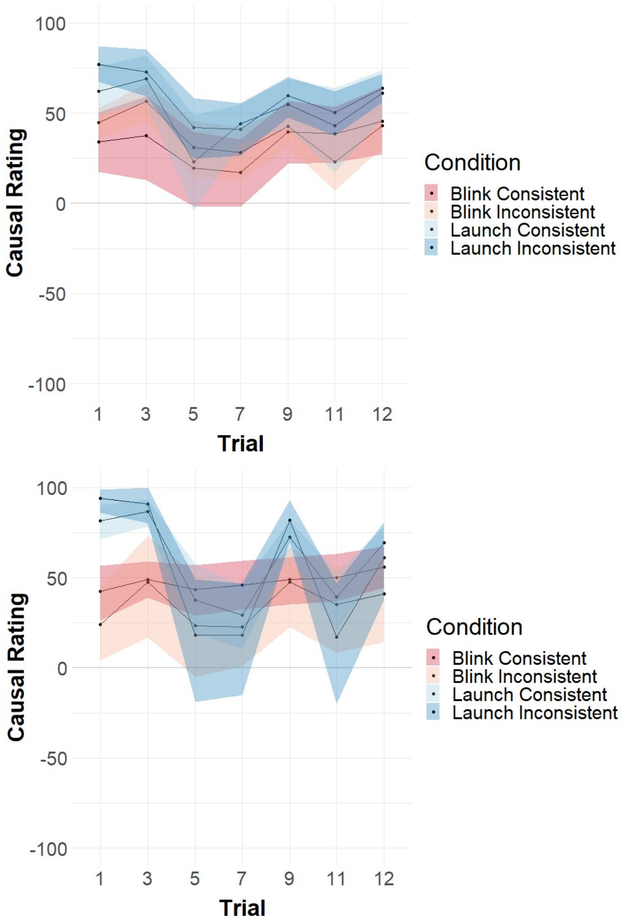

Figure 4: Mean causal ratings for probabilistic sequence for participants who passed (top) and failed (bottom) the manipulation recall check. Error ribbon: 95% CI.

For the probabilistic series, participants who failed the manipulation recall (Figure 4 bottom) also had a significant drop in causal ratings from trial 1 to trial 5 (B_5 = -44.0, SE = 8.44, t(136) = -5.21, p < .001, 95% CI = [-60.2, -27.8], $R^2_{partial}$ = .156), as well as lower *blink* than *launch* causal ratings at trial 1 (B_{blink} = -39.2, SE = 11.1, t(177) = -3.53, p < .001, 95% CI = [-60.4, -17.9], $R^2_{partial}$ = .146). There were also multiple interaction effects. The causal rating decrease was smaller for *blink* than *launch* from trial 1 to trial 5 ($B_{\text{blink},5}$ = 45.0, SE = 13.3, t(136) = 3.37, p < .001, 95% CI = [19.4, 70.6], $R^2_{partial}$ = .157), and also from trial 1 to trial 12 ($B_{\text{blink},12}$ = 25.8, SE = 13.3, t(136) = 1.94, p = .055, 95% CI = [-53.0, 11.7], $R^2_{partial}$ = .069). The violation-of-expectation effect at trial 5 was marginally greater when the instruction was inconsistent with the data ($B_{5,\text{inconsistent}}$ = -32.0, SE = 16.9, t(177) = -1.54, p = .060, 95% CI = [-64.4, 0.367], $R^2_{partial}$ = .015).

Proportion Ratings Due to limited space, we omit the graphs of proportion ratings as they are qualitatively the same as Figures 3 and 4; for completeness, the graphs are provided at https://osf.io/f3bjq/. In the deterministic series, participants who passed the recall check gave lower proportion ratings for *blink* than *launch* (B_{blink} = -18.2, SE = 6.61, t(186) = -2.76, p = .0064, 95% CI = [-31.0, -5.45], $R^2_{partial}$ = .102). When the initial instruction indicated (falsely) that the objects were unreliable, then ratings increased more for *blink* than *launch* at trial 5 ($B_{\text{blink},5,\text{inconsistent}}$ = 20.3, SE = 8.79, t(208) = 2.31, p = .0217, 95% CI = [3.35, 37.3], $R^2_{partial}$ = .025), and trial 12 ($B_{\text{blink},12,\text{inconsistent}}$ = 23.9, SE = 8.79, t(208) = 2.72, p = .0072, 95% CI = [6.90, 40.9], $R^2_{partial}$ = .034).

In the deterministic series, those who failed the recall check similarly gave lower proportion ratings for *blink* than *launch* (B_{blink} = -31.3, SE = 10.2, t(68.9) = -3.07, p = .0031, 95% CI = [-50.8, -11.8], $R^2_{partial}$ = .151). Proportion ratings were also lower at trial 1 when the instruction falsely stated that the objects were unreliable ($B_{\text{inconsistent}}$ = -22.2, SE = 10.4, t(68.9) = -2.13, p = .0372, 95% CI = [-42.1, -2.24], $R^2_{partial}$ = .065).

In the probabilistic series, participants who passed the recall test gave lower proportion ratings for *blink* than *launch* at trial 1 (B_{blink} = -26.8, SE = 10.1, t(194) = -2.66, p = .0084, 95% CI = [-46.2, -7.40], $R^2_{partial}$ = .065). Ratings decreased from trial 1 to trial 5 in response to the first failure case of the series (B_5 = -27.4, SE = 8.74, t(178) = -3.14, p = .0020, 95% CI = [-44.2, -10.6], $R^2_{partial}$ = .101), but this drop was smaller for *blink* than *launch* ($B_{\text{blink},5}$ = 28.0, SE = 10.7, t(178) = 2.17, p = .0096, 95% CI = [7.38, 48.6], $R^2_{partial}$ =.043).

In the probabilistic series, participants who failed the recall test gave lower proportion ratings for *blink* than *launch* at trial 1 (B_{blink} = -15.9, SE = 7.66, t(140) = -2.07, p = 0.0403, $R^2_{partial}$ = .111). Proportion ratings dropped significantly from trial 1 to trial 5, the first failure (B_5 = -10.9, SE = 4.94, t(136) = -2.21, p = 0.0286, $R^2_{partial}$ = .074), and the drop was larger when the initial instruction was inconsistent with the data ($B_{5,\text{inconsistent}}$ = -25.6, SE = 9.88, t(136) = -2.59, p = .0107, $R^2_{partial}$ = .027). Proportion ratings at trial 1 were marginally lower for *blink* than *launch* given inconsistent manipulation ($B_{\text{blink},\text{inconsistent}}$ = -25.5, SE = 13.7, t(1340) = -1.86, p = .0645, $R^2_{partial}$ = .024). Finally, the decrease in proportion rating at trial 5 was smaller for *blink* than *launch*, but this difference was marginally greater if the instruction was inconsistent with the data ($B_{\text{blink},5,\text{inconsistent}}$ = 23.8, SE = 14.0, t(136) = 1.70, p = .0915, 95% CI = [-3.04, 50.5], $R^2_{partial}$ = .021).

Summary Overall, *blink* elicited lower causal and proportion ratings than *launch* for early cases. Manipulation of expectations of determinism had no consistent effect. In the deterministic series, participants who passed the recall check had an increase in *blink* causal and proportion ratings across trials when the instruction (falsely) said that the objects were unreliable, yet those who failed the recall check showed little change across trials. In the probabilistic series, participants who passed the recall check showed no effect of instruction consistency, but participants who failed the recall check generally had a greater violation-of-expectation after the first failure case (trial 5). For proportion ratings, this effect was moderated by the perceptual features of the cause.

Importantly, both instruction consistency and recall check (and, to a lesser extent, type of cause) ceased to have an effect on learning by trial 12 in the series. Given the task context of alien objects, learning seemingly came to be almost entirely determined by the statistical data, rather than initial expectations invoked by instruction or perceptual cues.

Discussion

A few observations about Experiment 3 were notable. First, in the deterministic series, mean causal and proportion ratings of *launch* never approached ceiling, even though those ratings were consistently near ceiling in the previous experiments. The task context (alien objects) might have invited a baseline expectation of probabilistic behavior such that even an ordinary cause of launching was perceived as less-than-full-strength. Even so, perceptual details impinged on causal and proportion ratings, at least at trial 1, as seen by the main effect of type of cause.

Second, the manipulation of determinism expectation did not have a consistent effect across conditions. We found no effect with instruction consistency in the probabilistic series when participants passed the recall check, for both causal and proportion ratings. Although null results are hard to interpret, this lends support to the conjecture that the task context induced such a strong nondeterminism expectation that an explicit instruction was redundant. Notably, the initial impact of instruction consistency largely disappeared by trial 12. That is, a written manipulation of expectation of determinism swayed early causal judgments, but this impact faded out as participants received increasing numbers of observations.

Third, the statistical contribution of the manipulation of determinism expectation did not replace that of type of cause (*blink* vs. *launch*), regardless of performance on the recall check. In other words, written instructions did not explain away the effect of perceptual cues associated with the different causes of launching. In one possibility, the written manipulation was too brief: participants might need input about why these objects behave as they do (e.g., mechanism information). In another possibility, the alien object context induced so strong an expectation of nondeterminism that it rendered the written instruction less (but not completely in-) effective. In a third possibility, written instructions and perceptual cues are distinct contributors to participants' expectation of determinism such that manipulating one need not affect the other. Separating these possibilities requires further experimentation, but Experiment 3 implies that the impacts of type of cause (i.e., the perceptual features of physical causal events) and of verbal manipulation of expectations of determinism are not reducible to each other.

General Discussion

Past research offered many algorithms for incremental causal learning in which outcomes changed gradually with new statistical data. Yet human causal learning could also be abrupt, especially when people assume a causal relation to be deterministic (e.g., Michotte, 1963). In the present research, we examined how expectations of determinism interact with statistical input to shape causal learning in the context of physical causation. Different perceptual event features accompanied different expectations of determinism (Exp. 1A & 1B), and a violation of determinism expectation resulted in greater changes in causal learning outcomes than a violation of probabilistic expectation (Exp. 2). Notably, causal learning outcomes were affected by a brief written manipulation of their determinism expectation, but this effect occurred only early in the series and did not screen off the effect of perceptual cues (Exp. 3). Results showed that perceptual features affect how adults integrate and evaluate statistical data during causal learning. However, the relationship between perceptual features, determinism expectation, and causal learning remains a question for further research.

The present research leaves room for further inquiry. First, the written manipulation of determinism expectation in Experiment 3 was short and focused largely on the statistical properties of the alien objects. As previously discussed, it might take a much stronger manipulation to supplant the effect of perceptual cues, if that is even possible. Indeed, Schlottmann and Shanks (1992) provide "anecdata" that it is very difficult to override the effect of perceptual cues during causal learning, at least with physical causation.

Second, the present research leaves open the question of how to conceptualize the notion of determinism. Some have suggested that it functions as a continuous variable akin to causal strength (Lu et al., 2008; Yeung & Griffiths, 2015). Others have posited that people's notion of determinism is a categorical variable with two "modes": deterministic and nondeterministic (Yin & Sun, 2021). The sharp difference in causal and proportion ratings between *launch* and *blink* (Experiments 1A, 1B, and 2) is more consistent with a categorical notion of determinism, but this needs to be tested with other types of physical causation and across domains.

Overall, the present research showed that different expectations of determinism—as invoked by different perceptual event features—accompany different causal learning outcomes in a sequential learning task. A violation of the determinism expectation results in a much more drastic change in learning outcome compared to that of the nondeterminism expectation: such a pattern is not easily explained by any of the incremental models of causal learning. Furthermore, the effect of perceptual cues on causal learning is powerful and not easily supplanted by written manipulations of determinism. Future research on human causal learning should account for expectations of determinism and the different sources for these expectations.

Acknowledgments

The present research was supported by Carnegie Mellon University's GSA/Provost GuSH Grant funding. We also thank David Rakison for his valuable feedback during the design and data collection process.

References

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.

Bramley, N. R., Dayan, P., Griffiths, T. L., & Lagnado, D. A. (2017). Formalizing Neurath's ship: Approximate algorithms for online causal learning. *Psychological Review*, *124*(3), 301.

Cheng, P. W. (1997). From covariation to causation: a causal power theory. *Psychological Review*, *104*(2), 367.

Danks, D., & Schwartz, S. (2006). Effects of causal strength on learning from biased sequences. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 28, No. 28).

Deverett, B., & Kemp, C. (2012). Learning deterministic causal networks from observational data. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 34, No. 34).

Fernbach, P. M., & Sloman, S. A. (2009). Causal learning with local computation. *Journal of Experimental Psychology, 35*(3), 678-693.

Frosch, C. A., & Johnson-Laird, P. N. (2011). Is everyday causation deterministic or probabilistic?. *Acta psychologica*, *137*(3), 280-291.

Lu, H., Yuille, A. L., Liljeholm, M., Cheng, P. W., & Holyoak, K. J. (2008). Bayesian generic priors for causal learning. *Psychological Review, 115*(4), 955-984.

Michotte, A. (1963). *La perception de la causalité*, 1946. Trans. Tim and Elaine Miles. London: Methuen.

Schlottmann, A., & Shanks, D. R. (1992). Evidence for a distinction between judged and perceived causality. *The Quarterly Journal of Experimental Psychology*, *44*(2), 321-342.

Scholl, B. J., & Nakayama, K. (2002). Causal capture: Contextual effects on the perception of collision events. *Psychological Science*, *13*(6), 493-498.

Schulz, L. E., & Sommerville, J. (2006). God does not play dice: Causal determinism and preschoolers' causal inferences. *Child Development*, *77*(2), 427-442.

Strickland, B., Silver, I., & Keil, F. C. (2017). The texture of causal construals: Domain-specific biases shape causal inferences from discourse. *Memory & Cognition*, *45*(3), 442-455.

Yeung, S., & Griffiths, T. L. (2015). Identifying expectations about the strength of causal relationships. *Cognitive Psychology*, *76*, 1-29.

Yin, P., & Sun, J. (2021). Is causation deterministic or probabilistic? A critique of Frosch and Johnson-Laird (2011). *Journal of Cognitive Psychology*, *33*(8), 899-918.