Title
Expectations of Determinism Underlie Domain Effects on Adult Causal Learning

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

Authors
Dinh, Phuong (Phoebe) Ngoc
Danks, David

Publication Date
2023

Peer reviewed

Expectations of Determinism Underlie Domain Effects on Adult Causal Learning

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

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

Abstract

Prior knowledge can affect causal judgments by inducing expectations in learners. One important type of prior knowledge is about domains; for example, physical systems are typically believed to be more deterministic than psychological ones. We examine the role of these types of determinism expectations in shaping subsequent causal judgments, and we argue that they mediate effects of abstract domain-wide expectations. Study 1 shows that expectations of determinism—in particular, of sufficiency—positively predict judgments of strength and sufficiency ratings, even when holding statistical data constant. Study 2 shows that abstract expectations at the domain level affect causal judgments, but these effects are mediated (differently between conditions) by vignette-level expectations of determinism. These results jointly suggest a productive way to conceptualize domain effects on causal judgments, namely as effects of expectations about the causal relations being investigated.

Keywords: causal learning; determinism; prior expectation, sequential learning; domain; mediation

Introduction

Adults rarely enter the task of causal learning as a blank slate: their prior knowledge induces expectations that shape the kinds of inputs they seek, integration of those inputs, segmentation of a continuous flow of experience into causal chunks, and more (Hagmayer & Waldmann, 2002; Johnson & Keil, 2014; Sloman & Lagnado, 2004). A notable kind of expectation in human causal learning is **determinism**. Adults conceptualize determinism as the level of either sufficiency or necessity of a cause for its effect (Mayrhofer & Waldmann, 2016). Moreover, differences in determinism expectations have been shown to affect later causal judgments: when initial data suggest deterministic causation in a variant of the blicket detector task, both adults and children were more confident in the existence of a causal relation in later learning (Griffiths et al., 2011). Yet prior research often used categorical learning outcomes (e.g., intervention choice, directional prediction) which constrain our ability to describe the link between determinism expectations and causal learning. And even when continuous measures of learning were used, the experiments focused on sharply contrasting expectations of determinism, which limited what we can infer from learning outcomes (Dinh & Danks, 2022).

One potential reason for variation in determinism expectations is the domain of causation. People hold varying expectations of determinism for causal relations in the world. Physical causal relations are expected to be more deterministic than those in biology, psychology, or the social domain (Yeung & Griffiths, 2015; Yin & Sun, 2021). Similarly, causation in physics is expected to be more linear than in psychology or biology (Strickland, Silver, & Keil, 2017), and explanations for physical causation less complex than for psychological causation (Johnson, Valenti, & Keil, 2019). Even so, little is known about if and how these expectations manifest in causal learning.

The present research used three domains—physics, biology, and psychology as characterized in the intuitive theory literature on North American participants (Gopnik & Wellman, 2012)—to induce different determinism expectations and explore their relationship to causal learning outcomes from sequences of data. We hypothesized that higher expectations of determinism would correlate with smaller changes in causal judgments when the data agree with expectations (holding statistics constant). We thus expected a greater violation-of-expectation effect when participants encountered data contrary to expectations (Study 1). We also examined the pathways by which domain impinges on later causal learning: in particular, we hypothesized that expectations at the level of specific causal relations mediated these "domain effects" on causal judgments (Study 2).

Study 1

Method

Study 1 used a sequential learning paradigm to examine how determinism expectations affect causal learning outcomes.

Participants Adults (N = 240, 118 women, 114 men, 4 non-binary people, 2 genderqueer people) located in the US were recruited via Prolific Academic**.** 20 participants were Black or African American, 19 Asian or Asian American, 12 Hispanic, 3 Latinx, 1 Indigenous American, 14 multiracial, 167 white, and 4 who did not report their race or ethnicity. On average, participants finished the task within an expected timeframe (M_{time} = 17.0 min, SD_{time} = 8.09 min).

Stimuli Six vignettes were adapted from Johnson et al. (2019) to build the learning series, two in each domain: physics, biology, and psychology. Each vignette included a one to two-sentence description of two events at a generic level: one event (the putative cause) preceded another (the

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society.*

putative effect) without any explicit mention of causation. Physical vignettes were 1) ultraviolet waves undergoing some effect and showing irregular feedback, and 2) excess filtration leading to hyper-pressure in sewage systems. Biological vignettes were 3) soil being depleted of organic material and turning dry afterward, and 4) meals of a particular substance leading to iron deficiency later. Psychology vignettes were 5) people having a kind of anxiety and developing a dark sense of humor, and 6) children who experience erratic punishment bully others.

Measures of Expectations After reading the vignette assigned to them (e.g., "Some ultraviolet waves are found to undergo Planck's effect. These ultraviolet waves then show irregular feedback."), participants answered four measures of expectation. First, participants identified the cause and effect in short-response answers (using "NA" if they thought there was no causal)relation. Participants then rated the extent to which the putative cause made the effect happen (vignette strength expectation) on a 201-point scale (-100 = *The cause strongly prevented the effect*, 100 = *The cause strongly generated the effect*). They also estimated the number of contributing causes to the effect on a [0, 100] scale as a measure of vignette complexity expectation.

Sufficiency and necessity are key dimensions of adults' notion of determinism (Mayrhofer & Waldmann, 2016), and so we used measures of both. Participants predicted the number of hypothetical cases out of 100 in which the effect would happen if the cause was present (vignette sufficiency), and if the cause was absent (vignette necessity) using a [0,100] scale. Lower necessity expectation ratings indicate higher degree of actual necessity of the cause.

Participants also reported expectations ([0,100] scales) of (i) how strong causes tend to be (general domain strength); and (ii) the extent to which current events are determined by previous events (general domain determinism) for physics, biology, and psychology at a general level.

Measures of Causal Learning Outcomes After reporting vignette expectations, participants saw a series of 12 trials built from that vignette. After each trial, they gave three measures of causal judgment, given all the trials seen in the series. Participants rated the extent to which the cause made the effect happen on a 201-point scale (-100 = *The cause strongly prevented the effect*, 100 = *The cause strongly generated the effect*) and gave sufficiency and necessity ratings (both on a [0,100] sliding scale). These cumulative ratings mirrored the vignette expectation measures and tracked participants' learning outcomes across the series.

Design The study used a 3 (domain) × 2 (relation) × 2 (success) between-subjects factorial design; participants were randomly recruited into a condition. In each condition, participants first read the vignette assigned to them and provided vignette-level expectations of strength, sufficiency, and necessity. They then completed a 12-trial learning series; each trial presented an individual case of the purported causal relation portrayed in the vignette they just saw. In the 100% series, the cause preceded the effect in 12 out of 12 trials. In the 75% series, the cause did so only 9 out of 12 trials: "failure" cases in the 75% series always occurred on trials 5, 7, and 11 for ease of comparison across conditions. After each trial, participants gave three cumulative judgments—strength, sufficiency, and necessity—given all they had seen. Finally, participants completed two measures of general domain expectations of strength and determinism for physics, biology, and psychology; this order prevented participants from guessing that domain was the key independent variable.

Procedure After signing up on Prolific, participants were directed to Qualtrics to complete the study. Those who provided consent then participated in the study as described in the Design. Finally, they answered an instructional manipulation check to ensure proper attention, read an explanation of the study, and gave qualitative feedback about the study before being redirected to Prolific. Participants were compensated US $2.50 regardless of performance.

Analysis

General domain expectations and causal ratings were analyzed with linear mixed-effect regression (LME) via the R packages lme4 (Bates et al., 2015) and lmerTest (Kuznetsova, Brockhoff, & Christensen, 2017). Domain—a three-level nominal variable—was dummy-coded into two indicators. A random intercept for each participant was added. For vignette expectations, we used linear regression (LM) from the base R package stats (R Core Team, 2020). 95% CI were obtained with bootstrapping (R = 5000) by the R package boot (Canty & Ripley, 2022) for LM and an lme4 bootstrapping feature for LME. Semi-partial pseudo-R^2 for fixed effects were calculated via the r2glmm package (Jaeger, 2017) using the Nakagawa and Schielzeth approach (2013).

Results

General Domain Expectations Causation in physics was expected to be stronger (B_{psy} = -3.92, SE = 1.87, $t(478)$ = -2.09, p = .0368, 95% CI = [-7.53, -0.35], R^2 = .004) and more deterministic (B_{psy} = -4.92, SE = 1.74, $t(478)$ = -2.83, p = .00487, 95% CI = [-8.27, -1.61], R^2 = .007) than psychology.

Vignette-Level Expectations No domain difference in vignette strength expectations was significant (Figure 1). Sufficiency expectations were higher in physics than in psychology (B_{psy} = -13.98, SE = 4.43, $t(237)$ = -3.16, p = .00181, 95% CI = [-22.76, -5.26], R^2 = .040). Expectations of complexity were lower for physics than for biology (B_{bio} = 16.83, SE = 3.85, $t(237)$ = 4.38, p < .001, 95% CI = [9.19, 24.56], R^2 = .075) and psychology (B_{psy} = 24.94, SE = 3.85, $t(237)$ = 6.49, p < .001, 95% CI = [17.47, 32.47], R^2 = .151). Biology vignettes were expected to be less complex than psychology ones (B_{psy} = 8.11, SE = 3.85, $t(237)$ = 2.11, p = .0359, 95% CI = [0.786, 15.50], R^2 = .018).

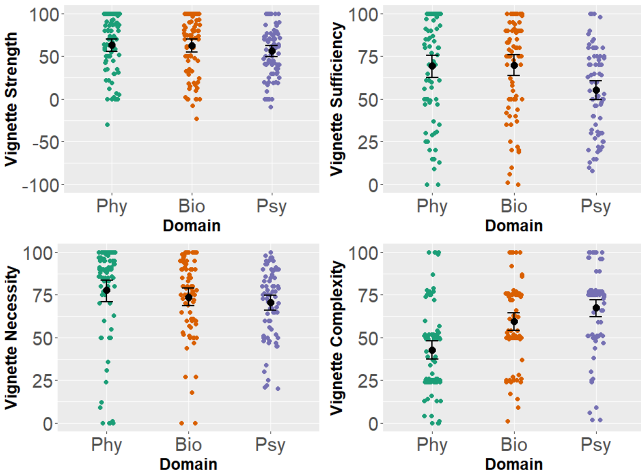

Figure 1: Vignette expectations of strength, sufficiency, necessity (reverse-scored), and complexity (Study 1).

Causal Learning Outcomes Analyses of causal learning outcomes focused on Trials 1, 5, and 12 (dummy-coded into two indicators); see Figure 2. An LME model was fitted for each type of causal judgment. Domain and trial number were interactive predictors; vignette-level sufficiency, necessity, and complexity expectations were additive predictors. Data for the 100% and 75% series were analyzed separately.

In the 100% series, strength ratings increased significantly from Trial 1 to 5 (B_5 = 15.18, *SE* = 4.47, $t(234)$ = 3.40, p < .001, 95% *CI* = [6.35, 23.80], R^2 = .015) and 12 (B_{12} = 19.03, *SE* = 4.47, $t(234)$ = 4.26, p < .001, 95% *CI* = [10.31, 27.55], R^2 = .023). Vignette sufficiency expectations predicted strength ratings in learning (B_{suff} = 0.506, *SE* = .0812, $t(114)$ = 6.23, p < .001, 95% *CI* = [0.347, 0.663], R^2 = .184).

Sufficiency ratings rose from trial 1 to 12 (B_{12} = 11.45, *SE* = 3.15, $t(234)$ = 3.63, p < .001, 95% *CI* = [5.30, 17.47], R^2 = .013). Vignette sufficiency expectations predicted sufficiency ratings (B_{suff} = 0.403, *SE* = 0.067, $t(114)$ = 6.02, p < .001, 95% *CI* = [0.272, 0.534], R^2 = .186). Sufficiency ratings were lower for psychology than for physics (B_{psy} = -14.81, *SE* = 5.61, $t(176.91)$ = -2.64, p = .009, 95% *CI* = [-25.68, -4.17], R^2 = .021). The increase in sufficiency ratings from trial 1 to trial 5 was also greater for psychology than for physics ($B_{psy:5}$ = 9.20, *SE* = 4.46, $t(234)$ = 2.06, p = .040, 95% *CI* = [0.448, 18.16], R^2 = .004). A similar pattern was found at trial 12 ($B_{psy:12}$ = 12.33, *SE* = 4.46, $t(234)$ = 2.77, p = .006, 95% *CI* = [3.35, 20.96], R^2 = .008).

Vignette necessity expectations predicted necessity ratings during learning (B_{nece} = 0.354, *SE* = 0.063, $t(114)$ = 5.61, p < .001, 95% *CI* = [0.229, 0.477], R^2 = .156). No other predictors reached significance.

In the 75% series, strength ratings dropped significantly from trials 1 to the first failure case at 5 (B_5 = -19.30, *SE* = 6.46, $t(234)$ = -2.99, p = .0031, 95% *CI* = [-32.06, -6.83], R^2 = .019). Vignette sufficiency expectations also predicted necessity learning (B_{suff} = 0.599, *SE* = .0816, $t(114)$ = 7.34, p < .001, 95% *CI* = [0.440, 0.759], R^2 = .181).

Sufficiency ratings dropped from trials 1 to 5 (B_5 = -13.60, *SE* = 3.73, $t(234)$ = -3.64, p < .001, 95% *CI* = [-20.98, -6.39], R^2 = .021). Vignette sufficiency expectations positively predicted sufficiency ratings during learning (B_{suff} = 0.550, *SE* = 0.0593, $t(114)$ = 9.26, p < .001, 95% *CI* = [0.437, 0.664], R^2 = .303). The drop in sufficiency ratings from trial 1 to trial 5 was greater for physics than for biology ($B_{bio:5}$ = 13.93, *SE* = 5.28, $t(234)$ = 2.64, p = .009, 95% *CI* = [3.93, 24.55], R^2 = .011) and psychology ($B_{psy:5}$ = 15.88, *SE* = 5.28, $t(234)$ = 3.007, p = .003, 95% *CI* = [5.51, 26.48], R^2 = .015). Similarly, the drop in sufficiency ratings from trial 1 to trial 12 was greater for physics than for biology ($B_{bio:12}$ = 14.63, *SE* = 5.28, $t(234)$ = 2.77, p = .006, 95% *CI* = [4.43, 24.84], R^2 = .013) and psychology ($B_{psy:12}$ = 15.27, *SE* = 5.28, $t(234)$ = 2.89, p = .00417, 95% *CI* = [4.65, 25.501], R^2 = .014).

Lastly, vignette-level necessity ratings predicted necessity ratings in the learning series (B_{nece} = 0.482, *SE* = .0588, $t(114)$ = 8.201, p < .001, 95% *CI* = [0.369, 0.598], R^2 = .265). Furthermore, necessity ratings increased significantly from trial 1 to trial 5 (B_5 = 8.75, *SE* = 3.024, $t(234)$ = 2.89, p = .004, 95% *CI* = [2.78, 14.59], R^2 = .012).

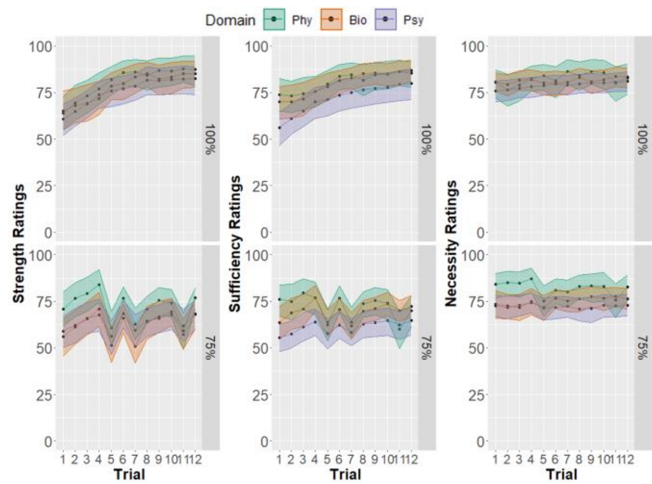

Figure 2: Strength, sufficiency, and necessity (reverse-scored) ratings in the 100% and 75% series (Study 1).

Discussion

Results supported our first hypothesis that vignette-level expectations of determinism—especially those involving sufficiency—shaped later causal learning outcomes. Higher vignette sufficiency expectations led to higher strength and sufficiency ratings regardless of the contingency data; the opposite was found for necessity expectations and necessity ratings during causal learning. The effects of these expectations on causal judgments were graded and small (in terms of raw coefficient, not pseudo-R^2) compared to that of the incoming data: on average, participants were most responsive to the trials that they encountered. Necessity judgments did not change as a function of trial number, which was sensible given that the cause was always present: these data provided no information about the necessity of the cause.

At first glance, results echoed prior research on domain expectations: adults in the US expected causation in physics to be stronger and more determined by prior events than in psychology, both at the general domain and vignette levels. Yet during learning, only sufficiency judgments showed any effect of domain. One possibility is that domain impacts strength and sufficiency judgments differently. Sufficiency measures tap into a statistical interpretation of causal strength but do not explicitly assert a causal relationship between two variables. As such, one may imagine a scenario in which C is sufficient for (in the predictive sense) but not causally related to E. Strength measures instead require participants to judge the causal link between C and E. A handful of cases may thus suffice for participants to update sufficiency judgments, but not be enough to reveal the nuances of domain effects on strength judgments. Another possibility is that prior expectations are themselves only "noisily" organized along domain lines, as closer investigation showed great variability in participants' vignette-level expectations. The design of Study 1 could not distinguish between these possibilities.

Study 2

Study 1 showed that prior determinism expectations shaped later causal judgments with statistical data, and that domain selectively informed causal judgment. Unfortunately, the between-subject design glossed over possible within-subject variability in expectations of causation across physics, biology, and psychology. The between-subject design also precluded examination of mediational links between domain, vignette-specific expectations, and causal judgments. As such, Study 2 borrowed the design of Study 1 but made domain a within-subjects factor.

Method

Participants Given the longer session compared to Study 1, we limited recruitment on Prolific to participants with ≥60 studies completed, a study approval rate of ≥90%, and physically located in the US. Adults (N = 195, 63 women, 127 men, 4 nonbinary people, 1 genderqueer person) participated in the study (M_{time} = 43.95 min, SD_{time} = 26.32 min). Of this sample, 19 were Black or African American, 13 Asian or Asian American, 4 Latinx, 8 Hispanic, 1 Indigenous American, 137 white, 11 biracial, and 2 unreported.

Stimuli In physics, the vignettes involved 1) explosions after some atomic particles were subjected to high heat and 2) underwater mountains having a smooth Eastern side following East-to-West sea currents. In biology, the vignettes were 3) a new bird species migrating to meadows and flowering plants increased subsequently, and 4) a new plant grew in some rivers followed by the increase of trout in those rivers. In psychology, the vignettes included 5) new parks being built, after which the wellbeing of nearby neighborhoods increased, and 6) workers feeling more satisfied at work after attending training on self-advocacy. We used different vignettes from Study 1 to check that our results were not due to the peculiarities of the vignettes.

Design and Measures Measures of expectations and causal learning outcomes were identical to Study 1. The vignette expectation task, sequential learning task, and general domain expectation questions made only two changes from Study 1. First, participants read three vignettes—one per domain—in random order and gave expectations of strength, sufficiency, and necessity for each. Second, participants saw three learning series, one per vignette. After reading vignettes and again after each series, participants rested for one minute by reading six puns unrelated to the task. Success (100% or 75%) was a between-participant factor.

Procedure The procedure was identical to Study 1, except as noted above. Participants were paid US $9.50.

Analysis

The analysis plan for general domain and vignette-level expectations was similar to Study 1. For causal judgments in the learning series, we used structural equation modeling (SEM) with the R package lavaan (Rosseel, 2012) to explore mediating effects of vignette-specific expectations on general domain expectations of strength and determinism.

Results

General Domain Expectations Participants expected causation in physics to be stronger than in biology (B_{bio} = -6.31, SE = 1.60, t(388) = -3.94, p < .001, 95% CI = [-9.47, -3.19], R^2 = .014) and psychology (B_{psy} = -13.6, SE = 1.60, t(388) = -8.49, p < .001, 95% CI = [-16.79, -10.47], R^2 = .062). Physical causation was also expected to be more deterministic than psychological causation (B_{psy} = -11.74, SE = 1.66, t(388) = -7.08, p < .001, 95% CI = [-14.95, -8.45], R^2 = .050). Lastly, biological causation was expected to be stronger (B_{psy} = -7.28, SE = 1.60, t(388) = -4.55, p < .001, 95% CI = [-10.39, -4.13], R^2 = .019) and more deterministic than psychology (B_{psy} = -9.21, SE = 1.66, t(388) = -5.55, p < .001, 95% CI = [-12.39, -6.028], R^2 = .031).

Vignette-Level Expectations An LME model was fitted for each type of expectation with a random intercept for participants. Predictors included domain, general domain expectations of strength and determinism, and their interaction with domain. General domain strength expectations positively predicted vignette lexpectations of strength ($B_{\text{Gen.Str}}$ = 0.280, SE = 0.117, t(563.40) = 2.38, p = .0175, 95% CI = [0.0509, 0.509], R^2 = .009) and negatively for vignette expectations of complexity ($B_{\text{Gen.Str}}$ = -0.294, SE = 0.101, t(563.56) = -2.91, p = .00373, 95% CI = [-0.490, -0.0972], R^2 = .014).

The positive association with vignette sufficiency expectations was marginally significant ($B_{\text{Gen.Str}}$ = 0.170, SE = 0.091, t(559.56) = 1.88, p = .061, 95% CI = [-0.0502, 0.3057], R^2 = .003). Lastly, higher general domain expectations of determinism also negatively predicted vignette-specific expectations of necessity ($B_{Gen.Det}$ = -0.143, *SE* = 0.0681, $t(521.69) = -2.094$, $p = .0368$, 95% *CI* = [-0.275, -0.00987], $R^2 = .006$).

The effect of general domain strength expectations on vignette complexity expectations was greater for physics than for biology ($B_{Gen.Det:Bio}$ = 0.334, *SE* = 0.136, $t(467.64) = 2.45$, $p = .0145$, 95% *CI* = [0.0689, 0.599], $R^2 = .009$) and psychology ($B_{Gen.Det:Psy}$ = 0.326, *SE* = 0.137, $t(504.81) = 2.37$, $p = .0181$, 95% *CI* = [0.0564, 0.594], $R^2 = .009$). No other interaction with domain was significant.

Mediation Analysis A SEM with linear regression was used to test the direct paths from general domain expectations of strength and determinism to each type of causal judgment in the learning series, as well as the indirect paths via vignette-level expectations of sufficiency and necessity. Trial number was added to each outcome model as covariates. The SEM fit for the 100% model ($\chi_2(5) = 13.14$, $p = .020$, CFI = 0.994, TLI = 0.961, RMSEA = 0.044, 90% *CI* = [0.016, 0.073], SMSR = 0.020) and 75% model ($\chi_2(5) = 36.55$, $p < .001$, CFI 0.975, TLI = 0.847, RMSEA = 0.085, 90% *CI* = [0.060, 0.112], SMSR = 0.031) were acceptable.

In the 100% series, no direct path from general domain expectations to any measure of causal judgment were significant (Figure 3). Vignette sufficiency expectations mediated the effect of general strength expectations on strength ($B = 0.102$, $p < .001$), sufficiency ($B = 0.062$, $p < .001$), and necessity ($B = -0.022$, $p = .007$). Vignette necessity expectations mediated the effect of domain expectations of determinism on strength ($B = 0.016$, $p = .036$) and necessity ratings ($B = -0.026$, $p = 0.018$).

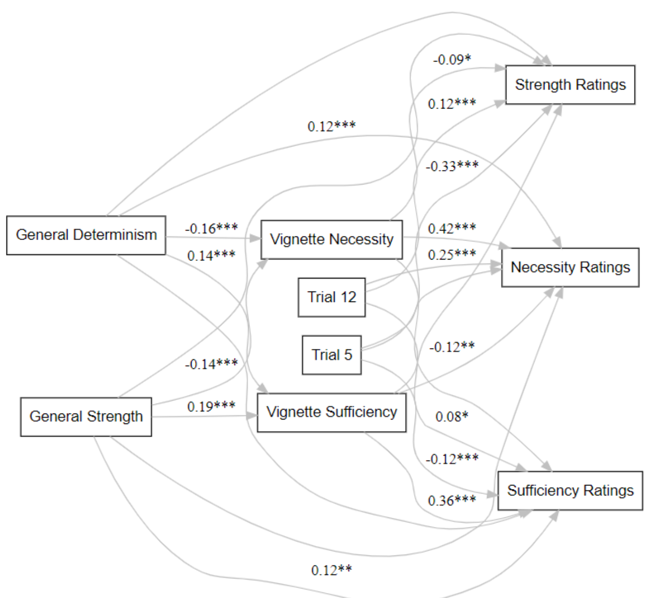

Figure 3: Structural equation model of the effects of general domain and vignette-specific expectations on 3 measures of causal judgments, 100% series (Study 2).

In the 75% series (Figure 4), general domain expectations of strength had a significant direct effect on sufficiency ratings ($B = 0.130$, $p = .003$). The direct path from general domain determinism expectations to necessity ratings was also significant ($B = 0.107$, $p < .001$). General domain expectations of strength had an indirect effect on strength ratings via vignette-level expectations of sufficiency ($B = 0.090$, $p < .001$) and vignette-level expectations of necessity ($B = 0.023$, $p = .038$). Furthermore, vignette-level sufficiency expectations mediated the effect of domain expectations of determinism on strength ratings ($B = 0.068$, $p = .004$). In addition, vignette-level expectations of sufficiency mediated the effects of general domain strength ($B = 0.075$, $p < .001$) and determinism ($B = 0.057$, $p = .002$) on sufficiency ratings. Lastly, all mediating effects of general domain expectations of strength ($B_{suff} = -0.019$, $p = .027$; $B_{nece} = -0.049$, $p = .001$) and determinism on necessity ratings were significant (B_{suff} = -0.014, $p = .035$; $B_{nece} = -0.059$, $p < .001$).

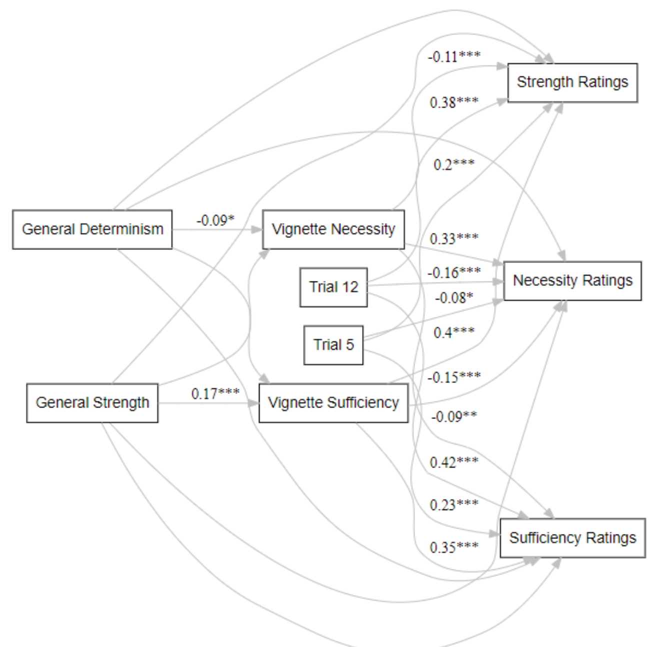

Figure 4: Structural equation model of the effects of general domain and vignette-specific expectations on 3 measures of causal judgments, 75% series (Study 2).

Discussion

Results about general domain expectations in Study 2 broadly echoed those of Study 1: on average, participants considered physical causation stronger and more determined by prior events than psychological causation, with biology fluctuating between them. Even so, there was considerable between-subject variability in participants' qualitative ordering of domain along those measures: for instance, a great number of participants thought causation in biology to be more deterministic than in physics. Furthermore, we found no significant main effect and very few interaction effects of domain on any vignette-level expectation when accounting for the continuous measures of general domain expectations. Measures of general domain expectations might capture more information about "domain effects" on vignette-level expectations than the domain labels themselves.

The SEM results provide some insights about not only the associations between different factors, but potential mechanisms to explain the results. In particular, vignette-level expectations mediated the effect of general domain expectations on most measures of causal judgments when given statistical data. That is, general domain expectations no longer provide predictive information once we know the vignette-level expectations. In the 100% data set, vignette-level expectations of sufficiency *fully* mediated the effect of general domain strength expectation on all three measures of causal judgments, while vignette-level expectations of necessity did so for general domain expectations of determinism. In the 75% data set, the mediation effect was partial for sufficiency and necessity ratings but full for strength ratings. Participants on average expected causation to be imperfect even at the general domain level: as such, the 75% success rate might have been interpreted as support for such early expectations. To our knowledge, this is the first empirical demonstration that the domain of causation is not "special" in and of itself for causal learning with statistical data, at least for a US adult sample. Rather, the expectations of determinism invoked by the specific causal relation being learned does most of the work of shaping learning outcomes.

General Discussion

Much research has shown how prior knowledge guide causal predictions, explanations, inference, judgments, and more: by inducing expectations in learners (e.g., Griffiths et al., 2011), or by influencing their choice of strategy for input integration (e.g., Waldmann, 2007). Our results supplemented those of Dinh and Danks (2022) to show that effects of determinism expectations—especially those of sufficiency—on causal judgments were persistent in the learning series though subtle compared to the effect of statistical data (Study 1). Higher vignette-level expectations of sufficiency accompanied higher strength and necessity ratings during learning, and higher vignette-level expectations of necessity did for necessity ratings during learning. This subtle effect of prior expectations may not apply to all participants, and the learning context was of low stakes such that participants might have felt comfortable trusting the data given to them. Future studies would benefit from a longer learning task, a wider range of data beyond statistical input, and a closer analysis of the individual patterns of learning outcomes. Still, current results reiterate a need to identify the sources of individual differences in human causal learning and highlight one such source, namely expectations of determinism.

Strikingly, vignette-level determinism expectations about specific causal relations mediated much if not all effects of abstract domain strength and determinism expectations on causal judgments. Any direct effect of general domain expectations only emerged when the data were congruent with these early expectations, namely that the cause works imperfectly (Study 2). As such, vignette-level expectations might offer a "mechanistic" account of how abstract expectations about certain kinds of causation manifest in judgments of concrete causal relations.

Another insight concerns the inconsistent role of domain—specifically, the categorical labels of physics, biology, and psychology—in vignette-specific expectations and causal judgments. First, the most consistent contrast in expectations at the general domain and vignette level is between physics and psychology, with biology hovering between the two (Study 1). The inclusion of general domain expectations of strength and determinism in the LME models for vignette expectations rendered most domain effects nonsignificant (Study 2). Second, domain effects were only present in sufficiency ratings during the series but not the other two measures (Study 1). Third, the structural equation models performed well with general domain expectations as exogenous variables.

Two possibilities about the status of domain effects on causal learning might be at play. In one possibility, "domain" is not a useful construct in explaining variance in causal learning. Rather, it is more productive to consider the expectations about particular causal relations or contexts that participants bring to the learning task. This amounts to a deflationary take on domain effects on causal learning and, more generally, causal cognition. This possibility has great merits given that domain is a highly fluid notion: for instance, the domain of animacy includes entities that has capacity for consciousness for adults in the US, but entities that has capacity to be in relation with others for Ngöbe adults in Panama (ojalehto et al., 2017). Even for participants within the same geographical region (which does not imply cultural hegemony), people vary greatly in expectations: there is a global pattern of physics being most deterministic, but a closer look suggests great between-subject variability.

Alternatively, domain effects exist but are heavily mediated and moderated by other effects pertaining to the specific learning stimuli and context: after all, vignette-level expectations did not fully explain the effects of abstract domain expectations on causal learning outcomes with statistical data in the 75% series (Study 2). Current studies cannot distinguish between these two possibilities, though results implicate accounts of not only causal learning but also conceptual organization (e.g., intuitive theory) and merits further study. If domain effects are present in some realms of causal cognition (e.g., expectations, categorization) but not others (e.g., causal learning), the status of domain in human cognition may not be as important as previously thought.

In sum, our studies showed that vignette-level expectations of determinism impinged on causal judgments (given statistical data) in considerable even if graded ways. These vignette-level expectations also mediated the effect of general domain expectations on causal judgments. Results open up new questions about the source and dynamic contributions of determinism expectations to causal learning.

Acknowledgments

The present research was funded by the Psi Chi Graduate Research Grant (Spring 2022), Carnegie Mellon University's GSA/Provost GuSH Grant, and Carnegie Mellon University's Center for Behavioral and Decision Research (CBDR) Graduate Student Small Research Grant Program. We thank Drs. David Rakison and David Plaut for their invaluable feedback on the experimental design of both studies. For Study 2, we thank the CBDR grant review panel for their helpful suggestions on the multi-part design of the study, as well as Dr. Jeanean B. Naqvi, Fiona Horner, and Kristina Dickman for their support with the Structural Equation Models.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. R package version 1.1-26.

Canty, A., & Ripley, B. D. (2022). *boot: Bootstrap R (S-Plus) Functions*. R package version 1.3-28.1.

Dinh, P. P. N., & Danks, D. (2022). Expectations of Causal Determinism in Causal Learning. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 44, No. 44).

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, *138*(6), 1085–1108.

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.

Hagmayer, Y., & Waldmann, M. R. (2002). How temporal assumptions influence causal judgments. *Memory & Cognition*, *30*, 1128-1137.

Byron Jaeger (2017). r2glmm: Computes R Squared for Mixed (Multilevel) Models. R package version 0.1.2.

Johnson, S. G. B., & Keil, F. C. (2014). Causal inference and the hierarchical structure of experience. *Journal of Experimental Psychology: General*, *143*(6), 2223–2241.

Johnson, S. G., Valenti, J. J., & Keil, F. C. (2019). Simplicity and complexity preferences in causal explanation: An opponent heuristic account. *Cognitive Psychology*, *113*, 101222.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13), 1-26.

Lagnado, D. A., & Sloman, S. (2004). The Advantage of Timely Intervention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(4), 856–876.

Lishinski, A. (2021). lavaanPlot: Path Diagrams for 'Lavaan' Models via 'DiagrammeR'. R package version 0.6.2.

Mayrhofer, R., & Waldmann, M. R. (2016). Sufficiency and necessity assumptions in causal structure induction. *Cognitive Science*, *40*(8), 2137-2150.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed effects models. *Methods in Ecology and Evolution*, 4(2), 133-142.

ojalehto, b. l., Medin, D. L., & García, S. G. (2017). Grounding principles for inferring agency: Two cultural perspectives. *Cognitive Psychology*, *95*, 50–78.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2), 1-36. R package version 0.6-13.

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

Waldmann, M. R. (2007). Combining versus analyzing multiple causes: How domain assumptions and task context affect integration rules. *Cognitive Science*, *31*(2), 233-256.

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.