Psychological Review

Rational Causal Induction From Events in Time

Tianwei Gong, M Pacer, Thomas L. Griffiths, Neil R. Bramley

Funded by:

- Air Force Research Laboratory
- NOMIS Foundation
- U.K. Research and Innovation, Economic and Social Research Council
- University of California Berkeley
- University of Edinburgh

Rational causal induction from events in time

Tianwei Gong *
Department of Psychology
University of Edinburgh

M Pacer *
Netflix, Inc

Thomas L. Griffiths
Departments of Psychology and Computer Science
Princeton University

Neil R. Bramley
Department of Psychology
University of Edinburgh

Author Note

* Denotes equal contribution. We thank Dave Lagnado, Maarten Speekenbrink, Marc Buehner, and James Greville for providing their stimuli and data (Greville & Buehner, 2007; Lagnado & Speekenbrink, 2010). This study was conducted under ethical approval granted by the Edinburgh University Psychology Research Ethics Committee (Ref No: 3231819/1). All data and analysis code are available at: https://github.com/tianweigong/rational_time. Support for this work was provided by a Edinburgh University PPLS scholarship to TG, a Berkeley fellowship and a National Defense Science & Engineering Graduate fellowship to MP, a Air Force Office of Scientific Research grant (FA9550-3-1-0170) and a NOMIS Foundation grant to TLG, and a EPSRC New Investigator grant (EP/T033967/1) to NB. Correspondence concerning this article should be addressed to Tianwei Gong, who is now in Department of Experimental Psychology, University College London, UK. E-mail: t-gong@ucl.ac.uk.

Abstract

A longstanding focus in the causal learning literature has been on inferring causal relations from contingencies, where these abstract away from time by collating independent instances or by aggregating over regularly demarcated trials. In contrast, individual causal learners encounter events in their daily lives that occur in a continuous temporal flow with no such demarcation. Consequently, the process of learning causal relationships in naturalistic environments is comparatively less understood. In this paper, we lay out a rational framework that foregrounds the role of time in causal learning. We work within the Bayesian rational analysis tradition, starting by considering how causal relations induce dependence between events in continuous time and how this can be modeled by stochastic processes from the Poisson–Gamma distribution family. We derive the qualitative signatures of causal influence, and the general computations needed to infer structure from temporal patterns. We show that this rational account can parsimoniously explain the human preference for causal models that invoke shorter, more reliable and more predictable causal influences. Furthermore, we show this provides a unifying explanation for human judgments across a wide variety of tasks in reanalysis of seven experimental datasets. We anticipate the framework will help researchers better understand the many manifestations of continuous-time causal learning across human cognition and the tasks that probe it, from explicit causal structure induction settings to implicit associative or reinforcement learning settings.

Keywords: causal induction; causal inference; continuous time; learning; Bayesian models

Rational causal induction from events in time

Time is inherent to our understanding of the world, shaping how we link the things that happen around us and the actions we take. We might judge that a backfiring car startled some birds, suspect that a new food gave us indigestion, infer from a boiling kettle that someone was recently in the kitchen, predict that we will be sore the day after the gym, or anticipate that a storm will follow a pink sunrise. All these inferences leverage causal models linking events in virtue of their experienced and historical temporal proximity through the lens of our intuitive causal theories.

The successes of everyday cognition, as well as the successes of our scientific theories and the technologies they support, suggest that people are capable of representing entities, properties, relations, events, states, and data defined in terms of time. This has been recognized since the earliest attempts by philosophers to define what it means to form beliefs about the external world (Hume, 1740). Time's arrow continues to be a core feature of philosophical discussion around the metaphysics of causality (Cartwright, 1994; Ross, 2024; Woodward, 2021), the acquisition of knowledge (Gettier, 1963; Goodman, 1983), and the functioning of intentional and volitional control (Dennett, 1971; Libet, 2009). It is not surprising then, that the study of human and animal learning has grown out of basic notions of association and reinforcement whereby the closeness of actions and events in time governs how we come to relate them in our minds (Gallistel et al., 2019; Gallistel & Gibbon, 2000; Garcia et al., 1966; Gershman, 2015; Hamou et al., 2025; Mnih et al., 2015; Rescorla & Wagner, 1972; Schultz et al., 1997; Tarpy & Sawabini, 1974).

In recent decades, cognitive psychologists used the approach of rational analysis (Anderson, 1990) to study how people learn causal structure from different kinds of environmental data (Griffiths & Tenenbaum, 2009). However, accounts of human causal learning have predominantly focused on inferences from contingency data. In these settings, evidence is provided helpfully "prepackaged" in the form of multiple (typically independent) trials or observations in which causal variables take different states (Allan, 1980; Anderson & Sheu, 1995; Cheng, 1997; Griffiths & Tenenbaum, 2005; Rescorla & Wagner, 1972). Consequently, the causal beliefs that emerge concern the probabilistic dependence between the states of causes and the states of effects, on average, without any representation of time. One common paradigm involves presenting participants with a set of independent samples in which putative causes and effects are either present or absent. Cover stories have been used to contextualize this as data arising from experimental research in biology (Buehner et al., 2003; Lu et al., 2008), physics (Coenen et al., 2015; Lagnado & Sloman, 2004), and psychology (Rottman & Keil, 2012), since multiple independent trials are often the data that scientists collect under laboratory conditions. A minimal example of this kind of task might involve pairs of patient outcomes (e.g., sick or not) under different treatment assignments (e.g., vaccinated or unvaccinated). Having seen some evidence, participants are asked to judge whether or to what extent the treatment causally affected the outcome (Buehner et al., 2003; Stephan et al., 2021). A 2-by-2 contingency table can capture the prevalence of different combinations of putative cause and effect states (Allan, 1980; Cheng, 1997), and where this indicates dependence there is evidence for some form of causal relation. Researchers have proposed a variety of approaches for drawing causal inferences from this sort of data and integrating new evidence with prior expectations (see Perales & Shanks, 2007, for a review) and distinguishing sharply between naturally observed and experimentally manipulated states (Lagnado & Sloman, 2002).

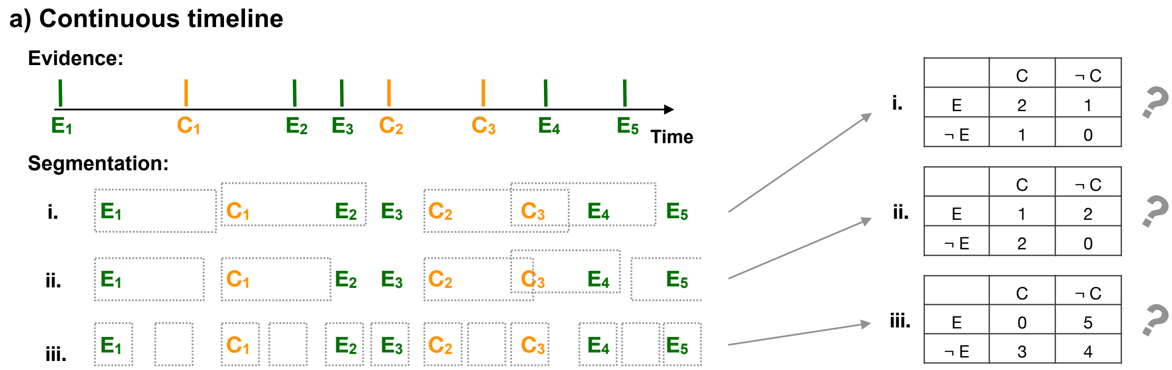
While these settings put timing considerations to one side, they do not eliminate them. Researchers in causal learning (Gong & Bramley, 2024; Greville & Buehner, 2007; Lagnado & Speekenbrink, 2010; Pacer, 2016) and associative learning (Gallistel et al., 2014; Gallistel & Gibbon, 2000; Hamou et al., 2025) have both recognized the problem of using "trials" as the basic unit of measurement. Fundamental questions remain as to how to determine an appropriate time window to measure outcomes, and how to ensure the observations are sufficiently independent to be aggregated. Without supporting knowledge about the relevant causal mechanisms, waiting too short a time before measuring an effect may not allow the influence to propagate or become apparent (e.g., the vaccine may not have taken effect yet), while waiting too long will tend to introduce confounding factors (e.g., the infection running its course, or the patient dying from natural causes). Equally, we need to determine the timing of interventions since some time-dependent factors (e.g., age) may also mediate the relationships between variables (Gong et al., 2023; Rottman, 2016). In order to curate scenarios and aggregate data into these simple contrasts, one must already draw on sophisticated prior causal beliefs about the relevant mechanisms and their temporal properties. Without this one could not be confident that an experimental protocol truly licenses abstraction to the level of contingency data. In short, time is integral to any general account of how we induce and represent causal models of our environment.

In this paper, we present a rational framework for causal induction from time. We lay out a computational-level treatment of the problem (Marr, 1982), building this up from basic principles of statistical dependence between events in time to formalize a grammar for continuous-time causal theories and a calculus for generating and comparing them with data. Our framework unifies the formalisms laid out in Griffiths and Tenenbaum (2009) and Pacer and Griffiths (2012, 2015) with those used in Bramley, Gerstenberg, Mayrhofer et al. (2018, 2019), Gong and Bramley (2023), Gong et al. (2023) (see also Bramley, Mayrhofer et al., 2017; Gong & Bramley, 2020, 2022; Stephan et al., 2020; Valentin et al., 2020, 2022). Many of the formal elements we use here appear in one or several of these papers. However, none of these papers unpack this into a general theory, nor generalize their modeling across a wide class of time-based causal inference settings. We here synthesize those works and for the first time formalize a general framework, demonstrating its underlying rationale, the derivation of core principles, and showcasing its broad scope and fit to behavioral data in a diverse array of tasks.

We situate our analysis within the Bayesian rational analysis tradition (Anderson, 1990), as this has proven very successful in developing a rational account of atemporal causal induction settings (Griffiths & Tenenbaum, 2005, 2009; Pearl, 2000; Rottman & Hastie, 2014). The main difference from these is that we link causal influence with dependence between *events* in *continuous time*, rather than their coincidence in independent trials (i.e., contingency). We show that this formalism anticipates and grounds the foundational principles of causal induction laid out by Hume (Hume, 1740) and enshrined in theories of associative learning (Gallistel et al., 2019; Gallistel & Gibbon, 2000; Rescorla, 1968). We show there is a natural bridge between the time-dependence and contingency level focus of established tools for causal inference (Bramley et al., 2015; Pearl, 2000).

As a rational, computational-level model, the computations involved demand assume accurate perception, infinite memory and computational resources. While assumptions are not aligned with the limitations of human learners they serve to describe the normative problem that heuristics and approximations should approximate (Anderson, 1990; Griffiths, 2020; Simon, 1982). We show how the normative analysis anticipates the *ceteris paribus* human preference for causal explanations that connect events via shorter, more reliable and more predictable causal influences. Furthermore, we present the first computational model that can provide a unifying explanation for human judgments across seven experimental datasets from the temporal causal learning literature.

Our analysis has deep connections with theories of time, rate, and conditioning, foundational to the animal learning literature (cf. Gallistel & Gibbon, 2000; Hamou et al., 2025). We will highlight connections with these throughout the paper, but here highlight a few ways in which we think our Bayesian treatment offers a novel and uniquely general perspective on these basic learning phenomena. Associative and reinforcement learning (whether model free or model based) are ultimately models of behavior, while Bayesian models describe the interplay of inductive biases and evidence in the formation of beliefs. Since causal models are by design, use-case-agnostic models of an agent's environment (Craik, 1967), it feels natural to conceptually separate the analysis of how agents form models, from analysis of how these models guide behavior, even if the conceptual distinction is often blurred in cognitive processing. An often implicit assumption of associative learning models is that we can rely on local link-by-link learning to build a global understanding. However this is in general a heuristic that will lead to causal misattributions (Btesh et al., in press; Fernbach & Sloman, 2009). By modeling causal structure induction at the rational level as selecting the globally most probable causal model, we can identify and explain these mistakes and biases as consequences of approximations rather than risking treating them as the right answer to the wrong question (Bramley, Dayan et al., 2017; Fernbach & Sloman, 2009; Griffiths & Tenenbaum, 2005, 2009; Pearl, 2000). Modeling structure induction as a model selection problem also helps in thinking about the imputation of hidden causes (Gershman et al., 2010; Gershman et al., 2015; Valentin et al., 2020). Bayesian learning models are also effective in describing setting in which people make choices across a much larger hypothesis space (Bramley, Dayan et al., 2017; Bramley & Xu, 2023; Griffiths & Tenenbaum, 2009); incorporating structured priors, including domain-specific causal theories and mechanistic knowledge of various kinds (Lu et al., 2008; Yeung & Griffiths, 2015); being sensitive to sample size (Griffiths & Tenenbaum, 2005); and providing uncertainty or confidence estimates (Kolvoort et al., 2025; O'Neill et al., 2022; O'Neill et al., 2024). We here focus on incorporating temporal information into the Bayesian framework, but readers may refer to the extensive body of previous research comparing Bayesian models with associative and reinforcement learning models more broadly (e.g., Courville et al., 2006; Fernando, 2013; Griffiths & Tenenbaum, 2005, 2009; Lake et al., 2017; Perales & Shanks, 2007; Tenenbaum et al., 2006) and recent work about how the languages of Bayesian models and model-based reinforcement learning models could relate to one another (e.g., Eckstein & Collins, 2020; Gershman, 2015, 2017; Wang, 2021). Nevertheless, the Bayesian approach is not the only method, in principle, that can provide predictions for temporal causal learning tasks. Many of the high-level ideas in this paper are aligned with those championed by associative



b) Episodic evidence (Greville & Buehner, 2007)

Experimental Group

	Day 0	Day 1	Day 2	Day 3	Day 4	Day 5
1	C				E	
2	C		E			
3	C	E				
4	C			E		
5	C					E

Control Group

	Day 0	Day 1	Day 2	Day 3	Day 4	Day 5
6						E
7		E				
8						
9				E		
10						

Figure 1

Continuous time evidence. a) Examples of the overly arbitrary decisions one could make when segmenting continuous timeline evidence into contingency evidence, along with the corresponding contingency tables. b) Examples of episodic evidence adapted from Greville and Buehner (2007). In the experiment, participants assessed the impact of a treatment (C) on the survival of bacterial cultures, considering culture death as the outcome (E).

learning literature, particularly Gallistel (Gallistel et al., 2014, 2019; Gallistel & Gibbon, 2000; Gallistel & Shahan, 2024; Gallistel & Wilkes, 2016) and Rescorla's earlier work (Rescorla, 1968). Their models stem from animal associative learning paradigms, which are not identical to human causal structure learning tasks. For example, all causal learning studies we review in this paper ask participants to report their inner beliefs, rather than analyzing beliefs indirectly through actions. As such, we mainly highlight the higher-level similarities in the text while making a more detailed comparison between our Bayesian account and temporal associative learning models (Gallistel et al., 2014; Gallistel & Gibbon, 2000; Rescorla, 1968; Schultz et al., 1997) in the General Discussion.

Desiderata for a Rational Theory of Causal Induction from Time

We demonstrate here four key desiderata for a rational model of causal induction from time: 1) providing predictions outside trial-based data settings; 2) capturing reasoning about dynamics and feedback; 3) grounding our core temporal-causal intuitions; 4) generalizing across a wide range of temporal causal learning tasks.

Going beyond trials and contingencies

We experience our environment in a single continuous timeline, making timing considerations a ubiquitous aspect of inference. Data arriving in continuous time will generally involve causes and effects that occur neither simultaneously, nor sufficiently separated to allow for any principled segmentation into "independent, identically distributed" (i.i.d.) scientific samples. This implies we must be able to litigate between competing causal explanations linking multiple events, even as they occur and recur within a single ongoing data-stream. For example, an everyday causal inference problem is trying to identify the cause of a recurring stomach ache. As shown in Figure 1a, because any decision about how to cluster and aggregate over potential trigger events and sickness episodes is arbitrary, there is no unique or fully principled encoding of this continuous-time data into a contingency table. An analogous problem in associative learning would be the difficulty of scoring the trials when more than one unconditioned stimulus (the effect) occurs, or determining the boundary of trials when neither the conditioned stimulus nor the unconditioned stimulus occurs (Gallistel, 2021; Gallistel et al., 2014). This brings home that any analysis that focuses exclusively on a discretized trial structure cannot take full advantage of the available metric information about the continuous time that has passed. Worse, such a representation can result in different conclusions depending on one's choice of measurement window.

The key to dealing with this problem is not to create pseudo-trials, but rather to shift the representational focus to explicitly model causal influences, in terms of how they shape the *delays* between particular causal events and, relatedly, how they shape the *rates* at which events of different types occur (Gallistel & Gibbon, 2000).

Reasoning about dynamics and feedback

Many causal processes in the natural world are cyclic (Malthus, 1872), and people frequently report causal beliefs that include feedback loops when allowed to do so in experiments (Kim & Ahn, 2002; Nikolic & Lagnado, 2015; Rehder, 2017; Sloman et al., 1998).[1] Cyclic systems can involve both excitatory or inhibitory feedback, which can result in complex, periodic and chaotic behavior (Davis et al., 2020). For example, a cyclic system might exhibit events occurring in a repeated alternating fashion: e.g., a bidirectional relationship $A \leftrightarrow B$ could generate a sequence of events A,B,A,B,A, ..., while the same system plus an output component $A \leftrightarrow B \rightarrow C$ could produce a variety of temporal patterns depending on the relative delays and reliabilities of the individual connections ($A, B, C, A, B, C, ...$, but also $A, C, B, A, C, B, ...$, or $A, C, B, A, B, C, ...$). Recognizing, predicting, explaining or controlling the behavior of such cyclic causal systems is only possible if one properly represents the temporal dimension. Contingency data, at best, blurs this dimension and the Bayesian network formalism typically recruited for causal analyses represents causal structure as inherently acyclic (DAGs; Griffiths & Tenenbaum, 2009; Pearl, 2000; Rottman & Hastie, 2014).[2] In order to study how people reason about real time causal systems, we need a framework that is able to represent these dynamic and continuous features.

[1] Formally, a causal mechanism is cyclic if it contains a feedback loop such that a causal variable in the system has itself as a descendant (Pearl, 2000).

[2] Workarounds are sometimes used to model dynamic and cyclic structure with existing tools. For example, the dynamic Bayesian network "unrolls" a repeated temporal structure over equally spaced time steps (Dean & Kanazawa, 1989; Rottman & Keil, 2012; Valentin et al., 2022), where a chain graph can be used to model cyclic substructures with undirected edges in an otherwise directed causal network (Lauritzen & Richardson, 2002). However, these approaches impose significant constraints on representation. The former constrains the expression of temporal information to equally spaced discrete time points, allowing each type of event to occur only once at each time point, and implicitly modeling all causal influences as having the same latencies. Chain graphs do not represent dynamics of the causal feedback but only their equilibrium distribution. These limitations do not seem well matched to everyday causal reasoning where we may think that effects can occur at any moment, be separated by intervals of arbitrary and often variable length, and where the ability to anticipate *when* something will happen is likely to be important.

Grounding our core causal intuitions

There are many empirical findings regarding how people process temporal information to learn causal relationships. We here summarize common intuitions regarding delay information that people adopt when making causal judgments: short delays, reliable delays, and delay expectations. One of our goals in this paper is to explain these three intuitions from a rational perspective.

Intuition 1: "Contiguity" – (Relatively) shorter delays are more likely to be causal

Perhaps the most foundational result in human and animal learning is that strength of association between events depends on the delay between their presentations. This is the contiguity effect in associative learning in human and animals, where the association formed between two events decreases as the delay increases (see Gallistel et al., 2019; Schultz, 2015; Tarpy & Sawabini, 1974, for reviews). Similarly, people tend to make stronger causal attributions between events that occur close together than far apart, especially when they don't have specific knowledge of the mechanisms involved (Buehner & McGregor, 2006; Greville & Buehner, 2007, 2010, 2016; Lagnado & Sloman, 2006; Shanks & Dickinson, 1991; Shanks et al., 1989).

This effect shows up when different cause candidates are studied under a shared context. For example, it could be when there are competing causes in a system, people tend to attribute an effect to the cause more closely preceding the effect. Lagnado and Sloman (2006) found that when participants frequently observed events in the order $A - C - B$, they were more likely to consider C to be the cause of B rather than A, even though in some cases A and B co-occurred without C (see our later analysis of this dataset; see also Bramley, Gerstenberg, Mayrhofer et al., 2018). It could also be that when causes are learned in different trials, people give higher causal ratings in trials where they observe shorter inter-event delays ("fast causes") compared to trials with longer delays ("slow causes"; Buehner & May, 2003; Greville & Buehner, 2010, 2016; Shanks &

Dickinson, 1991; Shanks et al., 1989). Although short- and long-delay causes are presented in separate trials, typically some context was shared across trials making “slow causes” slower in a *relative* as well as an *absolute* sense. For example, researchers used the same observation duration for both short- and long-delay conditions and included the same density of baserate effect events (Greville & Buehner, 2010, 2016; Shanks et al., 1989). When the shared context is reduced, the contiguity effect may disappear. This has been named as a *time-scale invariance* property in animal learning research by Gallistel and colleagues (Gallistel & Gibbon, 2000; Gallistel & Shahan, 2024; Gallistel & Wilkes, 2016; Kalmbach et al., 2019). For example, in Gallistel and Shahan (2024), rats learned associations with delays up to 16 minutes, as long as the training was also scaled to be longer. Lagnado and Speekenbrink (2010), whose findings we will model later, also found that human participants drew similar conclusions about short- and long-delay causes when the total observation time of a trial was scaled to match the causal delays (i.e., the observation period for long-delay trials was proportionally longer than that of short-delay trials; see also Zhang & Rottman, 2024) and the baserate was matched accordingly (i.e., the baserate was lower for long-delay trials; see later analysis).[3]

Researchers have used process-level factors to explain the short-delay intuition, such as the idea that the longer the delay, the harder it is for the cause to be sustained in working memory long enough to become associated (Ahn et al., 1995; Buehner & May, 2003; Einhorn & Hogarth, 1986). However, this explanation does not reconcile the contiguity results with the time-scale invariance results, which can instead be naturally

[3] There are confounds in early studies to be considered when interpreting the contiguity results. For example, a free-operant procedure was often used where learners could decide when and how often to press a button to activate the cause. Participants were found to press less often when the causal delay was long (Buehner & May, 2003; Shanks & Dickinson, 1991; Shanks et al., 1989) (controlled in Greville and Buehner (2010) where a similar number of presses was found across conditions), which meant that participants tended to amass less evidence for long-delay causes. In some earlier studies, effects of later interventions were be masked if the effect of an early intervention had not yet been revealed (e.g., Shanks et al., 1989) (controlled in Buehner and May (2003) and Greville and Buehner (2010) where effects would never be masked), which would significantly impact the empirical causal strength of long-delay causes, as more ineffective interventions could have been made during a long intervention-effect interval.

reconciled within a rational Bayesian framework. In a later section, we will show that the short-delay intuition is rational when short- and long-delay causes are learned with shared contexts, i.e., when (1) causes compete within the same causal system; or (2) causes are learned in different systems with the same observation duration and baserate. We will also explain how the time-scale invariance property emerges when the shared context is eliminated.

Intuition 2: Reliable delays are more likely to be causal

People tend to make stronger causal attributions when the delays between a putative cause and effect are similar across repeated observations (Bramley, Gerstenberg, Mayrhofer et al., 2018; Gong et al., 2023; Greville & Buehner, 2010, 2016; Lagnado & Speekenbrink, 2010). Greville and Buehner (2010) provides an anecdote that can serve as an intuitive thought experiment: suppose you always encounter traffic lights that take a very long time to change during your commute to work. You've heard a rumor that flashing your car's headlamps might help because the traffic lights would respond to the flashing lights of emergency vehicles. Now, suppose you try this, and the traffic lights do indeed change after a consistent delay of around 10 seconds. Compare this to a situation where sometimes the lights change very quickly after your headlamp flash, while at other times they take much longer. In which situation would you be more likely to believe that flashing the headlamps actually causes the lights to change? Greville and Buehner (2010) indeed found that people give stronger causal ratings when the delays between a putative cause and effect are drawn from a narrower distribution (e.g., 4.5-7.5 s), as opposed to a wider distribution (e.g., 3-9 s) even when the average delay length is the same (i.e., 6 s). Bramley, Gerstenberg, Mayrhofer et al. (2018) asked participants to select between two causal structures based on episodic evidence with three types of event occurring in a consistent orders (e.g., $A - B - C$) but variable temporal delays. They found that people favored the "Chain" structure ($A \rightarrow B \rightarrow C$) when the delay between A and C was variable but the delay between B and C was more reliable, and preferred the "Fork" structure ($B \leftarrow A \rightarrow C$) when the delay between A and C was reliable but the delay between B and C was variable (see later analysis for this dataset).

Although a preference for reliable delays seems intuitive, it is challenging to explain under associative or reinforcement learning theories. For example, Greville and Buehner (2010) demonstrated that, under the assumption of temporal-discounting reinforcement learning (Chung, 1965; Myerson & Green, 1995), the expected sum of rewards for two varied action-reward pairs should be greater than that for two unvaried action-reward pairs, which, counterintuitively, would lead to a preference for unreliable delays. The reliable-delay intuition also cannot be explained by a simple difference in the learning rate or the time required to reach the asymptote, as empirically, the preference for reliable delays remains regardless of whether participants learned for 2 minutes or 4 minutes (Greville & Buehner, 2010). In contrast, we will demonstrate that our Bayesian model naturally captures this intuition, as well as its stability to data exposure manipulations.

Intuition 3: Delays that match causal expectations are more likely to be causal

People also tend to make stronger causal attributions when the delay between a putative cause and effect is consistent with their causal–mechanistic understanding of the situation at hand (Bramley, Gerstenberg, Mayrhofer et al., 2018; Buehner & McGregor, 2006; Gong & Bramley, 2023; Hagmayer & Waldmann, 2002; Stephan et al., 2020). For example, Buehner and McGregor (2006) found that participants assigned higher causal judgments to the insertion of a ball that turned on a light on a physical apparatus when the light came on after a few seconds, rather than instantly, if they were aware that it would take time for the ball to roll through the apparatus and reach the light switch (see also Buehner & May, 2004). Similar results were found in 4-7-year old children (Mendelson & Shultz, 1976; Schlottmann et al., 2013). Hagmayer and Waldmann (2002) found participants judged whether an insecticide prevents mosquitoes by comparing prevalence of mosquitoes in fields with and without the insecticide, but judged whether planting flowers

Table 1
Dataset features.

Name	Reference	Base Rate	Prevention	Cycle	Delay Prior
Continuous timeline, effect specified:					
Earthquake	Lagnado and Speekenbrink (2010)	✔	✗	✗	✗
Device: Prevention	Gong and Bramley (2023)	✔	✔	✗	✔
Continuous timeline, effect unspecified:					
Device: Active Learning	Gong et al. (2023)	✗	✗	✔	✔
Episodic evidence, effect specified:					
Bacteria	Greville and Buehner (2007)	✔	✔	✗	✗
Future Bacteria	Gong and Bramley (2024)	✔	✔	✗	✗
Episodic evidence, effect unspecified:					
Computer Virus	Lagnado and Sloman (2006)	✗	✗	✔	✗
Device: Chain or Fork	Bramley, Gerstenberg, Mayrhofer et al. (2018)	✗	✗	✗	✗

Note: Human data are from Experiment 2 in Lagnado and Speekenbrink (2010), Experiment 1a in Gong and Bramley (2023), Experiment 1 in Gong et al. (2023), Experiment 1 in Greville and Buehner (2007), Experiment 1 in Gong and Bramley (2024), Experiment 1 in Lagnado and Sloman (2006), and Experiment 3 and 4 in Bramley, Gerstenberg, Mayrhofer et al. (2018).

prevents mosquitoes based on whether the prevalence of mosquitoes was affected the year after the flowers were planted, presumably expecting that flowers would take longer to influence the insect population than insecticide. We will show how this influence of expectations fits neatly in the Bayesian rational analysis.

Providing causal judgment predictions for various learning tasks

The final desideratum for a rational theory is that it should be able to offer quantitative causal judgment predictions for a wide variety of temporal causal learning tasks. Across the literature, different tasks have manipulated a wide variety of dimensions from the number and nature of the causal events, how they are spaced, and what participants have to infer. In these tasks, judgment patterns cannot simply be accounted by one or two of the intuitions we highlight but require the complete set. One reason for the variety of causal learning tasks is that temporal evidence can be accumulated in different ways, depending on the context. Evidence might be collected within a single extended encounter with a causal system, where all events occur within a single timeline, as depicted in Figure 1a. For instance, in Lagnado and Speekenbrink (2010), participants observed a geological system for several minutes, tracking the occurrence of “seismic wave” events (potential causes) and earthquake events (the effect) unfolding over time. Alternatively, evidence might also be gathered from multiple independent causal systems of the same type, as shown in Figure 1b. For example, in the research conducted by Greville and Buehner (2007), participants observed the timing of the death of multiple separate bacteria culture samples (the effect) after receiving a particular treatment (the potential cause). Instead of having multiple cause and effect events within a single timeline, there is one cause event and one effect event each with its own timeline, with these independent samples aligned by their cause or treatment time. In the rest of the paper, we refer to the former situation as “continuous timeline”, while the latter is termed “episodic evidence”.

There is also a distinction based on whether the effect variables are specified. In some cases, the effect variables are specified, and participants are asked to diagnose which of several candidate variables are causing them. This is similar to the traditional associative learning task, in which different cues have different (positive or negative) associations with the target reward, and animals learn to assigns credit accordingly (Gallistel et al., 2019; Rescorla & Wagner, 1972). In other cases, participants are tasked with determining the existence of a connection between two or more variables and the causal direction of these connections.

We will model seven human datasets that we categorized into four groups based on the nature of the evidence (continuous or episodic) and whether the effect variables are specified, as shown in Table 1. These datasets also vary in other task dimensions, including: (1) base rate: whether the effect occurred without any endogenous causes; (2) prevention: whether preventative causal relationships were involved in the response options; (3) cycles: whether feedback loop relationships were involved in the response options; and (4) delay expectations: whether participants were informed or pre-trained about causal delays prior to the task. We will show that lay people’s judgments consistently align with the rational framework in these tasks in later modeling sections.

A Rational Model of Causal Induction from Time

We lay out our theory in two steps. The first step demonstrates the *qualitative* differences in temporal patterns when two variables are related versus when they are unrelated to each other. The second step demonstrates how we can *quantitatively* infer causal structures using temporal information. This logic is very similar to the theory of Causal Bayesian Networks (Pearl, 2000, see below) for atemporal causal induction, which specifies both how statistical independence serves as a qualitative way to determine which variables are related, and how parameterized graphical models and interventional calculus further assist in causal structure learning and downstream inference. To streamline the descriptions, we provide equations for generative causal induction in the main text, while equations incorporating preventative causation are largely relegated to Appendix A.

Independence in time

In grounding causal induction from contingencies, Pearl (2000) starts from a principle of statistical independence. If A and B are perfectly independent, that is if $P(B|A) = P(B)$, Pearl argues this is nearest we get to confidence that there is no causal relation between A and B.[4] If A and B are dependent, that is if $P(B|A) \neq P(B)$, it follows that there must be *some* causal explanation for this dependence. It could be because A causes B, because B causes A, or because A and B share a (potentially distant) common causal ancestor. Interventions allow one to rule between these possibilities, by statistically disconnecting the intervened-on variable from its normal causes, such that if B depends on the intervened-on occurrences of A, that is if $P(B|\text{do}(A)) \neq P(B)$, or $P(B|\text{do}(A)) \neq P(B|\text{do}(\neg A))$ we can be confident that A is genuine causal parent or ancestor of B. By combining data from a series of interventions one can thus identify causal structure among a set of relata (Eberhardt et al., 2012; Steyvers et al., 2003). In associative learning, Gallistel et al. (2014) has also argued that any effective cause should

[4] This is known as the assumption of faithfulness (Scheines, 1997), meaning roughly the assumption that there are no additional statistically invisible causal relationships.

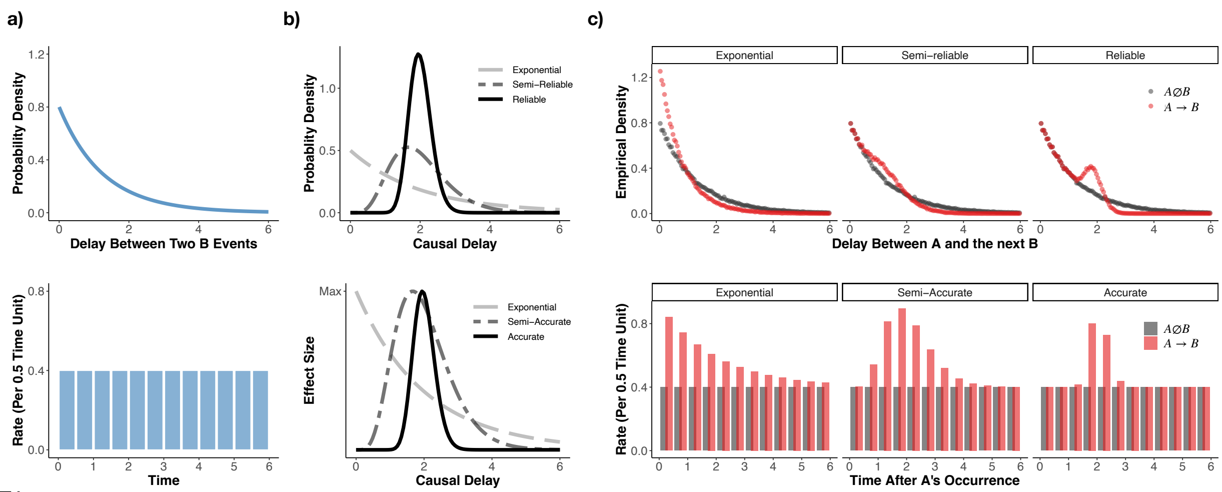


Figure 2

Causal dependence in time. a) The exponential distribution and constant rate are used to represent the base rate of B events. The same exponential distribution can be used to represent the delay between A and its next B when A and B have no relation. b) Different causal dynamics occur when the occurrence of A would generate either one extra B event (top) or a cluster of B events (bottom). c) How the data patterns differ when A causes B ($A \rightarrow B$) versus when A is not a cause of B ($A \emptyset B$).

provide more information about the effect's occurrence compared to a random time point. We will see later that our demonstration, especially under the special setting of an uninformative base rate (represented by exponential distributions), echoes this point.

Similar to Pearl (2000), we ground the temporal causal induction problem by first articulating the principle of independence with respect to temporal position before identifying causality with departures from independence. However, rather than having a simple probability of occurrence $P(B)$, we need to consider the *pattern* of B events occurring in time.[5] Without a model of its causes, a class of events might occur at subjectively unpredictable moments (e.g., receiving an email) or with some regularity or periodicity (e.g., receiving a repeat subscription delivery from Amazon). We here focus on the fully unpredictable cases for mathematical convenience, while our quantitative model is able to deal with periodic and otherwise more predictable event patterns (see our later

[5] We restrict our focus to the problem of inferring models relating events discretized as occurring at a point in continuous time. That is, we assume the learner starts having already processed their experience into *point events* that they are able to locate precisely in an experienced timeline. We discuss the relationships with other representations in the General Discussion.

analysis of the dataset from Gong & Bramley, 2023). If a type of events B occur completely unpredictably, i.e., the timing of the most recent B event provides no information about the timing of the next B event, it means the *delay* between any two adjacent B events will follow an exponential distribution (see Figure 2a, top panel). Exponential distributions are "memoryless" — their expectation is constant, and so does not depend on how much time has already elapsed (Gallistel & Wilkes, 2016; Gong, 2023; Grabenhorst et al., 2019; Pishro-Nik, 2014).[6] Under the exponential distribution, the expected *d*elay $P_d(.)$ from the present moment until the next event is always:

$$P_d(t|\lambda) = \lambda e^{-\lambda t} \tag{1}$$

The exponential distributions contain a *rate* parameter λ indicating how many B events one expects to observe on average; the delay between two adjacent B events is therefore $\frac{1}{\lambda}$ on average.

When the occurrence of a type of events follows an exponential distribution, the observed rate follows a Poisson process (see Figure 2a, bottom panel; Pishro-Nik, 2014). A Poisson process models the probability $P_r(k|\lambda)$ of observing a particular quantity k of such independent events in a fixed time unit given a presumed *r*ate λ:

$$P_r(k|\lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{2}$$

Here, λ is the same parameter shaping the delays between individual events in Equation 1. When delays are generated unpredictably at a constant rate λ, the Poisson process is *homogeneous.* The Poisson process, along with its corresponding exponential waiting time, has been widely used to model the arrival of events in previous cognitive research (Clarke, 1946; Gallistel & Wilkes, 2016; Grabenhorst et al., 2021; Grabenhorst et al., 2019; Griffiths

[6] Mathematically, this means if we expect to wait an average of x minutes for an event to occur, but we have already waited for a couple of minutes and the event as not happened yet, the expected wait time is still x minutes (see p.77 in Gong, 2023, for a proof).

& Tenenbaum, 2007). As we will see, maintaining these two representational perspectives (delays vs. rates) is very useful since the causal influence of events can be conceptualized as acting at either level, which can have subtle metaphysical and mechanistic consequences.

So far we have focused on the behavior of B while unperturbed by causal influences, i.e., the statistical patterns we expect to see between independent events. If A and B are independent (and do not share even a distant common cause), the occurrence of A cannot carry any information about the occurrence of B, so any inclusion of A in our model of B is predictively impotent. Instead of having a simple contingent probability of $P(B|A)$, we here need to specify the pattern of B conditioned on the occurrence(s) of A: if A has no influence on B, measuring the time from the occurrence of A until B occurs is equivalent to measuring from any other arbitrary moment before the next B event. Due to the memoryless feature, the delay from As occurrence to the next B will follow the same exponential distribution with the same parameter λ as that between one occurrence of B and the next (Figure 2a, 2c, top panel).[7] More intuitively, the rate of B occurrences after a causally impotent A will remain the same as the base rate (Figure 2a, 2c, bottom panel).

A qualitative understanding: Causal departures from independence

With a definition of independence in hand, we can start to articulate departures from independence and how they reveal causal structure. More specifically, we consider how the statistical pattern would look different, if occurrences of one type of event A are able to *generate* occurrences of another type of event B.

We here illustrate two different generative processes: one-cause–one-effect and one-cause–many-effects (cf. Gallistel & Wilkes, 2016), using a simple example in Figure 3. Suppose a fictional substance called 5-HTP is used to treat insomnia. Consuming a 5-HTP capsule can cause a person to sleep, resulting in a one-cause–one-effect scenario. Here,

[7] This may seem counter-intuitive: the fact that any A event inevitably occurs between two B events seems to suggest that the $A-B$ delay would be shorter than $B-B$ delay on average. However, to help build intuition, note that as long as A events are distributed independently from B, they are more likely to fall in a larger gap between successive B events, because these "take up more space" in the timeline.

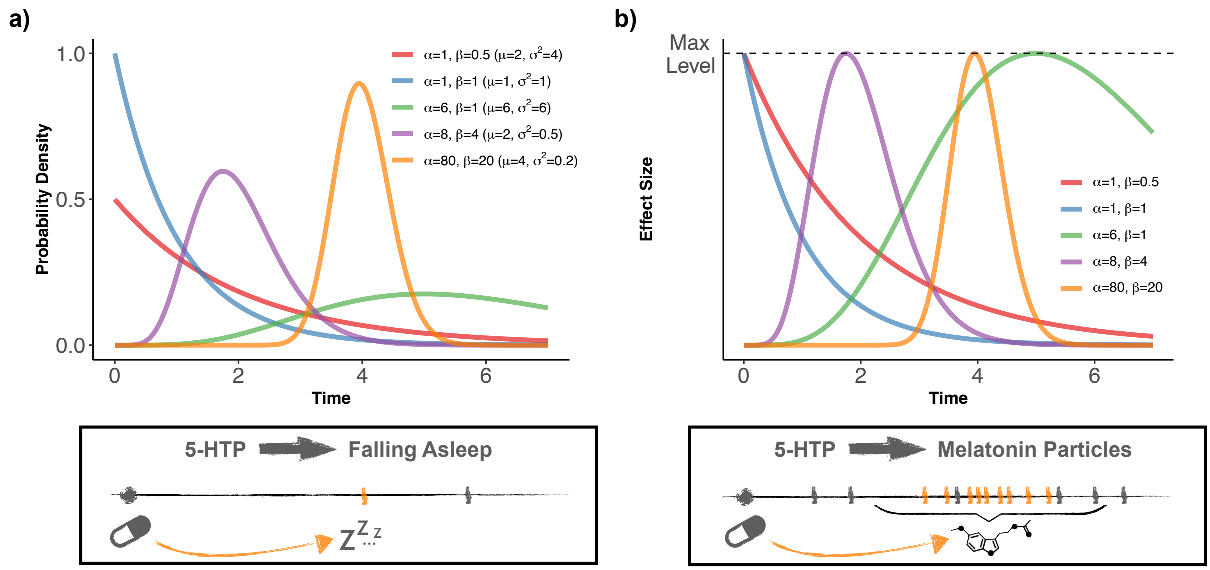


Figure 3

Examples of two types of function that could be used to model cause-effect delays and causal influences, respectively. Illustrative example relates a drug "5-HTP" and sleep. (a) A gamma probability density function capturing the delay between taking a drug and falling asleep and (b) scaled gamma density function capturing the rate of melatonin production after drug is administered. Different distributions demonstrate the functions' ability to capture various temporal dynamics. The orange distribution is the ground truth generative distribution. The orange effects in the timeline are those in fact generated by the drug while the gray effects are the base rate effects.

temporal information is embedded within the delay between the causal event of pill consumption and its effect event of falling asleep. The causal delay can vary across different mechanisms (see Figure 3a), analogous to our anticipation of certain medications (e.g., Adrenaline) taking effect rapidly and precisely, while others (e.g., painkillers) exhibit a delayed onset with some degree of variability. The gamma distribution can help us describe a variety of shapes and capture quick or slow temporal mechanisms. It is a generalization of the exponential distribution. It can be codified with a shape parameter α along with the rate parameter β:

$$P_d(t|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t} \tag{3}$$

With $\alpha = 1$ we recover the exponential, independent setting but with $\alpha > 1$, the gamma distribution becomes increasingly peaked and increasingly normally distributed around its mean, or expected delay (Figure 3a).[8] The expectation μ and the variance σ^2 of a gamma distribution follow $\mu = \frac{\alpha}{\beta}$ and $\sigma^2 = \frac{\alpha}{\beta^2}$.

It is also possible that one cause could generate many effects, for example, we might think the same scenario more granularly in terms of a pill's production of Melatonin particles over time (Figure 3b). One-cause–many-effect scenarios are prevalent in epidemiology. For example, a single water pollution event might cause many individuals to fall ill at different points in time (Griffiths & Tenenbaum, 2007). In this case, effects are generated at a rate level, which specify how many additional effect events we expect to be generated by the cause per time unit, and how this rate change is, itself, spread over time. This includes a functional form of the event's causal influence on the effect's rate over time, which may include an incubation period, peak, and a decay process (see Figure 3b). Concretely we can *also* use a scaled gamma function to capture the fluctuation in the rate of the effect, as a function of the time t since the cause happens and the peak rate λ_1 of the particular causal influence: $f(\lambda_1, t)$. This is done by scaling the Gamma density function via dividing by its mode, the density at $(\alpha - 1)/\beta$:

$$f(\lambda_1, t|\alpha, \beta) = \lambda_1 \cdot \frac{P_d(t|\alpha, \beta)}{P_d(\frac{\alpha-1}{\beta}|\alpha, \beta)} \tag{4}$$

After scaling, the predicted value ranges from 0 to the peak rate λ_1 (see Figure 3b).[9]

[8] To further build the intuition of what a gamma distribution models, it can be helpful to think of it as a sum of α exponential delays each of rate β. Here the role of β is equivalent to the rate parameter λ in the exponential distribution. A classic example of an unpredictable event is a radioactive decay of the type measured by a Geiger counter. If one estimated how long it would take a Geiger counter (placed near a source of radiation) to reach a count of α, the resulting delay distribution would be gamma with a shape of α and a β reflecting the average gap between each individual event. The larger α, the larger the mean but, relatively speaking, the narrower the spread of expected waiting times around that mean. Thus, a highly reliable delay is one that decomposes into a sum of many small independent unpredictable delays. This works similarly to how a normal distribution can be conceptualized as the sum of many independent errors with smaller errors producing a narrower distribution.

[9] We assume this distribution for convenience, but in principle any function with $(0, \infty)$ support can play this role. For instance, some causal influences might exhibit a step function, or remain at their peak level for an extended period before decaying, or be succeeded by a rebound effect. We will discuss these kinds

From a cognizer's point of view, the temporal dynamics are liable to be *variable* and uncertain. This is an inevitable feature of any model that abstracts away some of the detail, leaving unmodeled noise and complexity in the generative or measurement processes. The gamma family here helps to capture how abstract subjective probability distributions encode causal-model-based expectations about inter-event delays or event rates, and how these distributions can be shaped and sharpened with evidence, which provides a solid foundation for a rational model of time-based causal induction.

Figure 2b demonstrates three examples when the occurrence of A can result in either one (top panel) or multiple (bottom panel) events of B. In the single-effect scenario, the delay between A and its effect event B could follow a memoryless exponential distribution, a semi-reliable gamma distribution (with high variance), or a reliable gamma distribution (with low variance). Regardless of the situation, the delay between A and the next B would no longer resemble the scenario where A and B are independent, as depicted in Figure 2c. Similarly, in the multiple-effect scenario, the causal influence could manifest as an exponential shape (with no incubation time), a semi-accurate gamma shape (with a wide spread), or an accurate gamma shape (with a narrow spread). As shown in Figure 2c, the rate of B's occurrence will deviate from the constant rate when A is a cause of B. As such, the deviation in data patterns could help the learner obtain a qualitative understanding of whether A causes B or not.

A quantitative understanding: Structure inference from temporal information

So far we have demonstrated, at a qualitative level, how we can expect the temporal pattern of events following a putative cause event A to differ depending on whether A is truly a cause of B. We introduced two causal generative processes (generating a single-effect or multiple-effects), which inspired two analysis approaches: one based on the

situation in analyzing one of the datasets (Gong & Bramley, 2023). However, we believe utilizing the gamma function as a generic basis for modeling temporal causal beliefs is a sensible default, capable of capturing many scenarios from previous experiments (Gong et al., 2023; Greville & Buehner, 2007; Lagnado & Sloman, 2006; Lagnado & Speekenbrink, 2010). In other contexts the functional form can be derived from mechanism knowledge.

delays between particular token cause and effect events (event-based scheme), and one based on fluctuations in the rate of occurrence of the effect depending on occurrence of the cause (rate-based scheme). We will further elaborate the similarities and differences between these in this section. We will also show that event-based and rate-based approaches are not necessarily tied to the generative processes that inspire them. Rather, they are better understood as two ways of thinking.

A quantitative-level analysis is needed to formally infer causal structure. One can try to directly derive some indices from the qualitative patterns. Gallistel et al. (2014) pointed out the option of using the entropy between two effect distributions — measured from the previous cause event vs. any random point before (see Figure 2c) — to measure the associative strength. Although this seems sensible in pairwise association scenarios, it is not clear it can work for structure induction in general since it will change depending on the causal background, and because, unless it is an intervention, the information one variable carries about another can often be due to their sharing a common cause. We here follow the logic of inference to the most probable parameterized Causal Bayesian Network (Griffiths & Tenenbaum, 2005, 2009; Pearl, 2000; Rottman & Hastie, 2014) to build our quantitative approach, which will allow the model to both (1) infer the parameters capturing how a cause generates or prevents the effect in time; (2) deal with situations where multiple causes (or causes and base rates) *combine* with each other to influence the occurrences of an effect.

Challenging our initial 5-HTP example, everyday continuous-time evidence is often more complicated: events can occur at any time point, and different potential causes can overlap in time leading to pervasive credit assignment questions. For instance, imagine that you are taking a pill for medical purposes but feel that you are frequently experiencing stomach discomfort afterward. You might wonder whether this discomfort is genuinely a side effect of the pill. We illustrate this evidence in Figure 4a: one might experience multiple stomach aches during the observation period both related and unrelated to the medicine. Meanwhile, the pill may be consumed irregularly, and the stomach discomfort events caused by a pill, if they exist, could occur even after ingestion of a subsequent pill. Therefore, it is not possible to simply segment this kind of evidence into independent trials to compare candidate causal models.

The causal question can be formalized as whether treating the pills as an additional cause of stomach aches provides a better overall account of the evidence than treating the stomach aches as happening spontaneously (i.e., due to unexamined causes). Two hypothetical structures S_0 and S_1 are illustrated in Figure 4b. In S_0, only the base rate B causes the discomfort, while in S_1, both the base rate B and the pill taking C cause the discomfort. However, various other factors can potentially complicate this picture. For example, if the learner suspects that something else, such as their diet, may be contributing to their stomach aches, there could be additional dietary events in the timeline and we could incorporate their potential causal influence into the model comparison. We can imagine other cases, such as if you recognizes that your pill-taking is influenced by your stomach aches (e.g., if you avoid taking the pill when you already have a stomach ache), if the two could have a potential common cause such as time of day, or if the stomach aches could have their own feedback cycle that makes them occur with regularity. All these will refine the causal structure induction problem in ways our model framework is equipped to handle.

How can we address the structure selection question given temporal evidence? Three critical components have been highlighted for a rational account of causal induction (Griffiths & Tenenbaum, 2009): (1) an ontology that outlines the entities under investigation and their properties, (2) a set of plausible relations that suggest how entities may be connected, and (3) the functional form that determines how causes influence their effects under each type of relation. In the contingency setting, the ontology is a set of variables, the set of plausible relations is a hypothesis space of causal Bayesian networks and the functional form is often assumed to be noisy-OR combination of independent generative or preventative influences. Despite differences in the data they operate over, temporal and atemporal causal induction share the same basic problem of articulating model selection within a hypothesis space of causal structures. The normative learner updates their prior belief over structures s in the hypothesis space S using the likelihood function $P(\mathbf{d}|s;\mathbf{w})$ to arrive at a posterior distribution $P(s|\mathbf{d};\mathbf{w})$, given data $\mathbf{d}$ and a set of parameters $\mathbf{w}$:[10]

$$P(s|\mathbf{d}) \propto \int_{\mathbf{w}} P(\mathbf{d}|s;\mathbf{w}) \cdot P(s;\mathbf{w}) d\mathbf{w} \tag{5}$$

In the remainder of this section, we address the question of what constitutes an appropriate ontology, set of relations and functional form for the likelihood of causal event data in time $P(\mathbf{d}|s;\mathbf{w})$. Similar to the rule of qualitative patterns, we will demonstrate two approaches, with an event-based scheme that analyzes evidence at the individual delay level while a rate-based scheme analyzes the evidence at the rate level.

An event-based scheme

The event-based scheme we propose uses the concept of token-level "actual causation" to map each event to its possible causes (Halpern, 2016), identifying which of several candidate events actually caused the observed outcome (Gerstenberg et al., 2021; Stephan et al., 2020). While we may have knowledge and expectations about the delay between a cause and its effect (i.e., its mean and variance parameters), to derive these empirically we have to also commit to a particular causal story about which cause event actually produced which effect event in order to apply those expectations. Under this scheme one can consider various possible causal pathways that could produce the observed events, depending on the underlying causal mechanisms (Bramley, Gerstenberg, Mayrhofer et al., 2018; Gong & Bramley, 2023; Hamou et al., 2025; Stephan et al., 2020; Valentin

[10] Here we foreground the problem of structure selection rather than parameter estimation conditioned on a structure (Griffiths & Tenenbaum, 2005). That is, we assume that for each structure and functional form, the relevant parameters are theoretically marginalized over their prior and support if they are unknown. However the same mathematical formalism can straightforwardly be used for parameter estimation within a causal model.

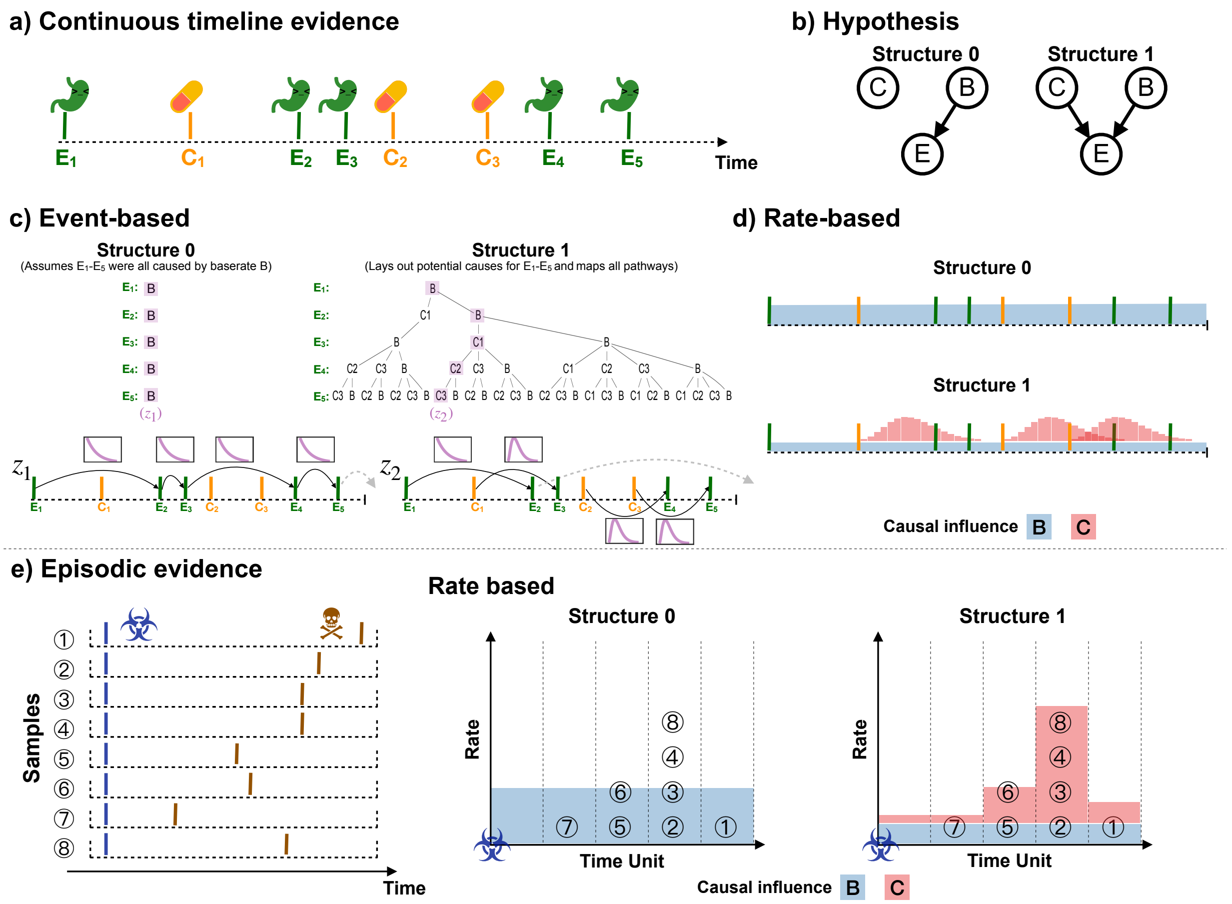


Figure 4

Causal inferences based on continuous-time causal evidence. (a) Evidence as events of stomach discomfort and pill taking unfolded in the timeline. (b) There are two causal structures in the hypothesis space. (c) The event-based scheme lays out all possible pathways (branches) that explain all effects under each hypothetical structure. (d) The rate-based scheme model in what way the rate of effects are expected to change under each hypothetical structure. (e) Episodic type of evidence where the cause and effect only happen once in each individual observation. Cases illustrate the situation in Greville and Buehner (2007) where the effect events across samples are assumed to follow exponential delays if the evaluated cause does not work. Under this situation, the evidence can be collapsed under the rate-based scheme.

et al., 2022). For example, in a causal structure s that includes an endogenous cause C, a hidden background cause B, and an effect E, each effect event could be caused by either C or B, resulting in a pathway set $\mathbb{Z}_s$ that contains a total of 2^k possible pathways (where k is the number of effects). The event-based scheme allows for specific mechanistic constraints to be integrated into pathway construction. For instance, if we observe a sequence of events, such as $\{C_1, E_1, E_2\}$, and also believe that this is the kind of system within which one C event can only cause one E event (Ross, 2024), we can rule out the pathway that assumes both E_1 and E_2 were caused by C_1.

Given that conditional on a structural hypothesis, the potential actual causal pathways are mutually exclusive and exhaustive, it follows that the overall likelihood of each structure hypothesis is the sum of the individual likelihood of these pathways:

$$P(\mathbf{d}|s;\mathbf{w}) = \sum_{\mathbf{z}\in\mathbb{Z}_s} P(\mathbf{z}|s;\mathbf{w}) \tag{6}$$

To determine the likelihood of a specific actual pathway given a hypothesized type level causal structure $P(\mathbf{z}|s;\mathbf{w})$, we compute the likelihood of both observed effect events e and unobserved but predicted events h. For each observed effect e, we evaluate the probability that it was caused by its presumed generative cause event g (denoted $g \rightarrow e$). Hidden (expected to be generated by a g but unobserved, denoted $g \rightarrow h$) effects contribute to the likelihood wherever we do not observe the expected effect of a generative cause. This could be due to (1) the generative cause failing to produce the effect or (2) the effect not having occurred yet:

$$\begin{aligned} P(\mathbf{z}|s;\mathbf{w}) = &\prod_{g\rightarrow e\in\mathbf{z}} \underbrace{w_g \cdot P_d(t_e - t_g|\alpha,\beta)}_{\text{Observed effects must have been generated}} \\ &\times \prod_{g\rightarrow h\in\mathbf{z}} \underbrace{(1-w_g) + w_g \cdot P_d(t_h > t_{end}|\alpha,\beta)}_{\text{Unobserved expected effects must have failed or be still-to-occur}} \end{aligned} \tag{7}$$

In Figure 4c, the event-based scheme generates pathways for explaining stomach discomfort under different structure hypotheses. For S_0, all effect events are attributed to the base rate. For S_1, any effect event could be attributed to the base rate or to cause events that occurred previously.

A rate-based scheme

The rate-based scheme models causes that temporarily affect the rate of occurrence of some effect. For a generative cause like 5-HTP, we expect the rate of its effect (melatonin) to temporarily increase from its base rate, and intuitively expect such rate increases to be additive (unless there are also interactions between the base rate causes and the focal cause). That is, an independent generative cause is something that adds extra events to the timeline without affecting those that would have been there anyway (Gallistel & Gibbon, 2000). For example, we might think of a large gathering causing rates of infection with the Covid-19 virus to spike by contributing additional infection events.

The rate-based scheme employs a non-homogeneous Poisson process (Pacer & Griffiths, 2012, 2015) to capture the likelihood of events in a setting where cause temporarily affect the rate at which their effects occur. The likelihood depends on how the observed rates at each time bin are aligned with the expected rates:

$$P(\mathbf{d}|s;\mathbf{w}) = \prod_t P_r(d_t|f(\lambda,t)) \tag{8}$$

We may be able to treat the base rate of an effect as a constant λ_0 if we have no information to suggest it is periodic or structured across time. For any generative cause, the causal influence can be modeled as modifying the effect's rate in a continuous fashion $f(\lambda_1, t)$. For example, an incubation-decay process is shown in Figure 3b, captured by the influence function in Equation 4.

Poisson processes have a desirable property known as "superposition", where the union of two independent Poisson processes with rates λ and λ' is still a Poisson process with rate $\lambda + \lambda'$. The superposition property not only give us a simple answer to the combination of a base rate and a (constant) causal influence, but also how a non-constant causal influence implies a fluctuating rate. Combining a set of generative causes $\mathbf{g}$ with a base rate of λ_0, the total expected effect rate $f(\lambda, t)$ at the time unit t can be represented by accounting for superposition as follows:

$$f(\lambda, t) = \lambda_0 + \sum_{i \in \mathbf{g}} f(\lambda_i, t) \tag{9}$$

This could be seen as a continuous-time version of the noisy-OR logic gate used in modeling contingency data (Cheng, 1997). Prevention can be similarly captured as filtering λ_0 resulting in a proportional temporary rate decrease that can be seen as a continuous-time version of the noisy-AND-NOT logic gate (see Appendix A).

Figure 4d illustrates how the rate-based scheme models causally-induced rate changes to explain stomach discomfort. In S_0, the model assigns a constant base rate to account for the number of effect events per unit of time. In S_1, the model incorporates the assumption that the effect rate dynamically changes following the occurrence of a cause event.

Summary and comparisons of two schemes

We have introduced two schemes for thinking about causal induction in continuous time. An *event-based* scheme involves reasoning at the level of individual events adjudicating whether the delays between putative cause–effect pairs are causal (E_i was actually caused by C_i) or coincidental (E_i was actually caused by an unobserved exogenous factor modeled by its base rate). The *rate-based* scheme involves modeling whether and how the rate of occurrence of an effect changes after a cause occurs. These schemes represent two different approaches to thinking about the temporal evidence, but there is a continuity between the two built on a shared mathematical foundation. As we will see below, there are cases where both perspectives coincide exactly or approximately in their predictions, regardless of the generative model of the evidence. Meanwhile, there are also cases where they differ in their predictions or computational cost. The event-based scheme can integrate detailed mechanistic knowledge, such as if cause produces exactly one effect, while it often demands a costly marginalization over different causal pathways consistent with evidence. The rate-based scheme implicitly marginalizes over these possibilities allowing it to deal with larger event counts but as the cost of making it hard to accommodate certain mechanistic principles.

Human Generic Delay Intuitions: Short, Reliable, and Expected

We highlighted that people see short, predictable, and expected delays as cues to causal connection. We now demonstrate that these intuitions all fall out from a Bayesian analysis under our model framework. We first provide intuitive explanations for each phenomenon and then back each up with simulation results. Each intuition found in humans can manifest in two types of tasks: *structure induction* and *causal diagnosis*. For example, the preference for the reliable delay can refer to learners' (1) higher causal ratings for an evaluated causal relation in trials where the delays are more consistent, compared to when they are less consistent, and (2) preference for Cause A over Cause B as an explanation for the effect in a single learning trial, when the delay between A and the effect is more consistent than the delay between B and the effect. This distinction matters more for the short-delay intuition than the other two, as we will discuss in the intuitive explanation section. In the simulation section, we primarily focus on structure induction and demonstrate how the same intuition readily applies to diagnostic causal reasoning.

Intuitive explanations

Delay length. In diagnostic causal reasoning (attributing the effect E to A over B in the single learning trial), even in the absence of specific causal delay expectations, a Bayesian learner has a preference for the diagnosis that implies shorter causal delays for at least two reasons. The first reason arises from the fact that causal delays can range from zero to infinity (with a limit on the lower side but not the upper side). As a result, any proper delay density will be right-skewed, inherently favoring smaller values. The second reason is that, when the learner don't have a specific prior for the delay variance, the possible range of the delay variance becomes larger as the delay mean increases, making long delays ceteris paribus harder to predict accurately. Figure 5 provides an example of

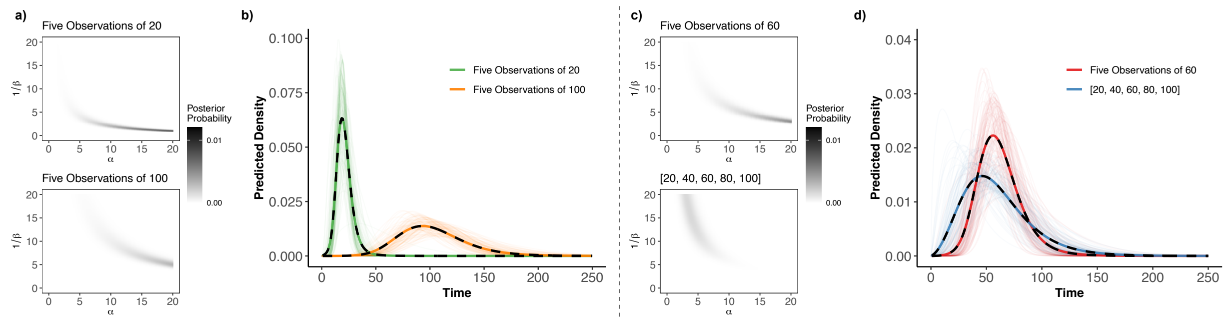


Figure 5

Parameter posteriors and posterior predictions after observing a 20-time-unit inter-event delay five times vs. 100-time-unit inter-event delay five times (a,b) and after observing a 60-time-unit inter-event delay five times vs. 20-, 40-, 60-, 80-, 100-time-unit inter-event delays (c,d). (a and c) Parameter posteriors using the conjugate prior update process under the assumption of an initially weak prior ($s = 10^{-6}, r = 10^{-6}, p = 1, q = 1$ for the distribution conjugate to the gamma likelihood with an intractable normalization constant; Fink, 1997). (b and d) The predicted density under sampled parameter combinations. Faint colored lines represent densities under posterior α and β samples. Thick and dashed black lines represent the marginal or posterior predictive distribution for future inter-event delays.

the posterior on parameters α and β (Figure 5a) and posterior predictive distribution after observing 20-time-unit delays five times vs. 100-time-unit delays five times (Figure 5b).[11] This illustrates that the posteriors are tighter after observing delays of 20 vs. 100. The delay prediction (Figure 5b) is also sharper (more confident) after observing delays of 20, in contrast to that following delays of 100. It highlights that a causal model that connects a set of events via shorter causal delays will assign higher likelihood to data than an alternative causal model that connects those events with longer causal delays, even when those delays are perfectly reliable (Greville & Buehner, 2010).

In structure induction (inferring the existence of short-delay or long-delay causes across different trials), the same preference appears because this comes down to a competition between the putative cause and the base rate to explain the effect's occurrences. The likelihood of the data under the baserate-only hypothesis (S_0 in Figure 4b) remains constant, while the likelihood the causal variable explains the occurrence depends on the length of the delays it implies, following the same logic

[11] We here used the analytical conjugate prior update process. In the rest of the paper, we will use simple Monte-Carlo sampling since analytic methods are not feasible.

mentioned in diagnostic causal reasoning. The posterior distribution between causal and baserate-only hypotheses, will differ accordingly. It is important to note though that if the base rate and causal delays are scaled up equally, and the prior on causal delays is uninformative, we will then have timescale invariance in the sense that a rational causal learner will favor the causal hypothesis equally in the “slow” and the “fast” dataset.[12] A straightforward way to demonstrate this is to generate long-delay stimuli by scaling short-delay stimuli by a constant (Gallistel & Shahan, 2024).

Delay reliability. The reliable-delay intuition can simply be explained by the *likelihood* calculation. If observed delays exhibit great variability, the resulting gamma distributions spread their expectations wider, resulting in a lower marginal likelihood compared to less variable delays. As in the delay length section, we provide an example in Figure 5 showing the parameter posteriors (Figure 5c) and delay predictions (Figure 5d) after observing a consistent 60-time-unit delay five times, versus observing a set of varied delays with a same average on 60 time units.

Delay variance would naturally be scaled with the delay length in a time-scale invariant environment. Otherwise, if the researchers force the long delays (e.g., 6 s) and short delays (e.g., 3 s) to have the same absolute standard deviation (e.g., 0.1 s), the relative variance will be smaller for the long delay, providing an advantage in the long-delay condition. We will further demonstrate this in the simulation section and when analyzing Lagnado and Speekenbrink (2010).

Delay expectation. The expected-delay principle can be understood as the influence of mechanistic knowledge or prior experience on people’s *prior* distribution regarding a causal delay. For instance, if individuals strongly believe that a genuine switch should take approximately 4 seconds to turn on a device, a switch that takes 2 or 6 seconds would have a lower prior probability and consequently a less good explanation for an effect

[12] However, a fully uniform prior over delays, e.g., $\exp(\frac{1}{\infty})$, is improper because the range of possible delays is infinite meaning the prior has zero density everywhere, necessitating use of Markov Chain methods or weakly informative priors in practice.

Table 2

Symbols used and their meanings under three contexts.

	Synthetic data	Event-based scheme	Rate-based scheme
m_u	Mean of causal delays.	–	–
i_u	Half interval of causal delays.	–	–
k_b	Number of base rate effect events.	–	–
k_c	Number of cause events.	–	–
w_c	Cause's success probability.	Cause's success probability.	–
μ	–	Mean of causal delays.	Mean of causal influence function.
σ^2	–	Variance of causal delays.	Variance of causal influence function.
μ_b	–	Mean of base rate delays.	–
σ_b^2	–	Variance of base rate delays.	–
λ_0	–	–	Effect's base rate.
λ_1	–	–	Max generative causal influence.
ξ	–	–	Max preventative causal influence.

Note: σ_b^2 is only used when modeling Gong and Bramley (2023) which included periodic base rates. In other cases, the base rate delay was modeled using exponential distributions, which only included one parameter.

compared to a switch that was pressed four second earlier. Correspondingly, a device activation 2 or 6 seconds after a switch press would be more likely to be caused by its base rate than a device activation 4 seconds after a switch press (Buehner & McGregor, 2006).

Simulation

To demonstrate that our account exhibits these features, we now simulate and model synthetic data. Given that most of the human empirical evidence was based on the one-cause–one-effect context, we focus on that context here. Note that the rate-based scheme generically assumes that a cause can produce multiple effect events and hence is slightly inconsistent with our simulation setting. Nevertheless we show that it still demonstrates a sensitivity to the change of delay lengths and delay variance. The meanings of symbols used in simulation and later dataset modeling are summarized in Table 2.

To show that this framework can handle data that are not exclusively generated from gamma distributions (just like humans, Greville & Buehner, 2010), we generate synthetic delay stimuli using uniform distributions, denoted as $U(l, u)$, with a lower bound l and an upper bound u. We used the following procedure:

1. Define each synthetic stimulus as lasting for 300 time units.

2. Generate k_b base rate effect events and place them at random times sampled from $U(0, 300)$.

3. Generate k_c cause events, again placing these on the timeline by sampling from $U(0, 300)$.

4. Generate effect events given the assumption that each cause event has a probability w_c of producing one effect event E with a delay sampled from $U(m_u - i_u, m_u + i_u)$. In other words, the ground truth for the simulated data is always the structure that includes the causal link as well as the base rate.

We then estimate the probability that there is a generative causal influence from C to E, i.e., judging the posterior probability of S_1 over S_0 in Figure 4b. The evidence that data $\mathbf{d}$ provide in favor of S_1 over S_0 can measured by the log-likelihood ratio:

$$\log \frac{P(\mathbf{d}|S_1; \mathbf{w})}{P(\mathbf{d}|S_0; \mathbf{w})} \tag{10}$$

We assume that all parameters used for simulating data are unknown to the learner and hence the models marginalize over the parameters $\mathbf{w}$. We achieve this via simple Monte Carlo integration drawing $m = 10,000$ prior samples for all parameters. For the event-based scheme, we assume a uniform prior on causal strength, meaning that cause succeeds in producing its effect with a probability $w_c \sim U(0, 1)$. We also assume weak priors on the parameters of the gamma causal delays between Cs and their effects E, such that they have mean $\mu \sim U(0, 300)$ and variance $\sigma^2 \sim U(0, \mu^2)$. This means the prior on the mean causal delay is equally likely to be anything between zero and the full length of the episode while the delay shape parameter ($\alpha = \frac{\mu}{\sigma^2}$) can be anything between 1 and ∞.[13]

[13] An alternative approach is to sample the mean (μ) and shape (α) parameters from very flat exponential distribution (Bramley, Gerstenberg, Mayrhofer et al., 2018). This produces the same qualitative results.

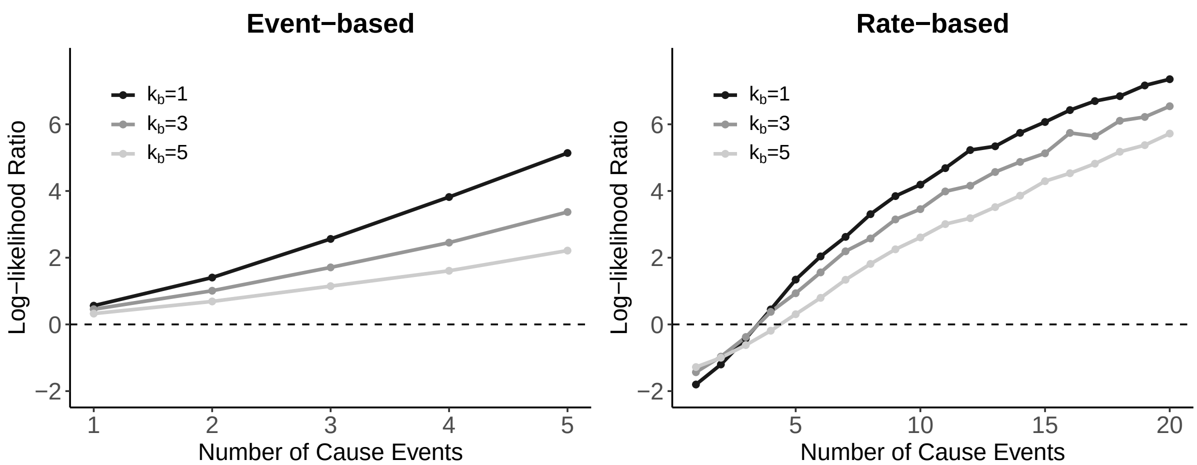


Figure 6

How log-likelihood ratio changes with the amount of cause events.

Similarly, the model assumes delays between base rate events follow an exponential distribution with mean $\mu_b \sim U(0, 300)$.

For the rate-based approach, we specify priors on the base rate $\lambda_0 \sim 1/U(0, 300)$, the max causal influence $\lambda_1 \sim U(0, 1)$. We assume causal influences dynamically inflate the rate according to a scaled gamma distribution with priors on the mean $\mu \sim U(0, 300)$ and the variance $\sigma^2 \sim U(0, \mu^2)$.

Synthetic data

We first implemented a sanity check task to make sure that the model's preference for S_1 over S_0 (1) increases as the number of cause events in the synthetic data increases and (2) decreases as the base rate of effect events increases. Both of these properties are based on the principle of distinguishing the (causal) signal from the noise and have been demonstrated in the atemporal causal learning setting (Cheng, 1997; Griffiths & Tenenbaum, 2005; Wu & Cheng, 1999). We here use $w_c = 1$, $m_u = 15$, $i_u = 5$, and consider different numbers for baserate events k_b ($k_b = \{1, 3, 5\}$). For the rate-based model, we consider numbers for cause events k_c ranging from 1 to 20 (with a step of 1), while for the event-based model, we limit the range to 1 to 5 due to computational cost. Results are shown in Figure 6. In both cases, the log-likelihood ratio in favor of S_1 increases as the

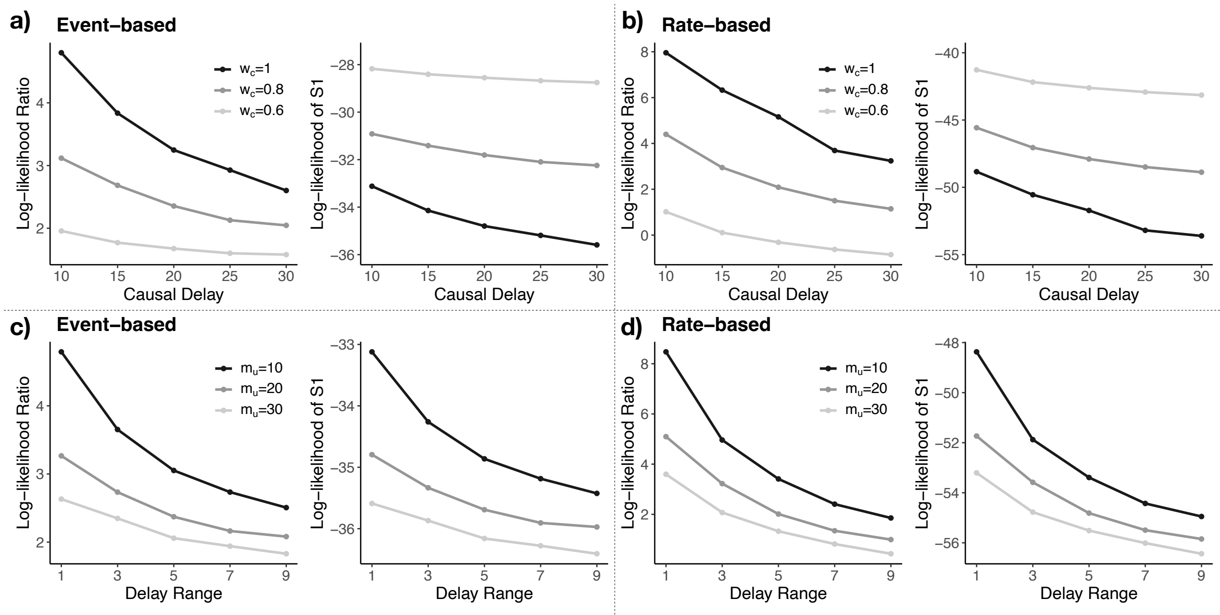


Figure 7

How log-likelihood ratio changes with the causal delays and delay ranges (and the corresponding log-likelihood of S_1). A ratio above zero indicates that the model favors S_1 (the causal structure) over S_0 (the base rate structure).

number of cause events increases. Both models perform better when the base rate is low. This reflects that the model's basic ability to learn causal structure from temporal data. Compared to the event-based model, the rate-based model requires more cause events to favor S_1 over S_0, indicating a higher requirement for data points under the relaxed constraints assumed by the model (it does not require specifying how many effect events would be generated by one cause event). Accordingly, we will use $k_c = 5$ for the event-based model and $k_c = 10$ for the rate-based model in the later simulations.

Delay length

To illustrate the short-delay preference found in humans, we simulate stimuli arranged in a grid with $w_c = \{0.6, 0.8, 1\}$, $m_u = \{10, 15, 20, 25, 30\}$, $i_u = 1$, and $k_b = 3$. Note that a fixed baserate events k_b over a fixed observation duration here is key to inducing this preference. Figure 7a demonstrates that the event-based model's preference for S_1 over S_0 diminishes as the duration of the true causal delays increase. This observation supports the notion that causal attribution is stronger when the delay is shorter

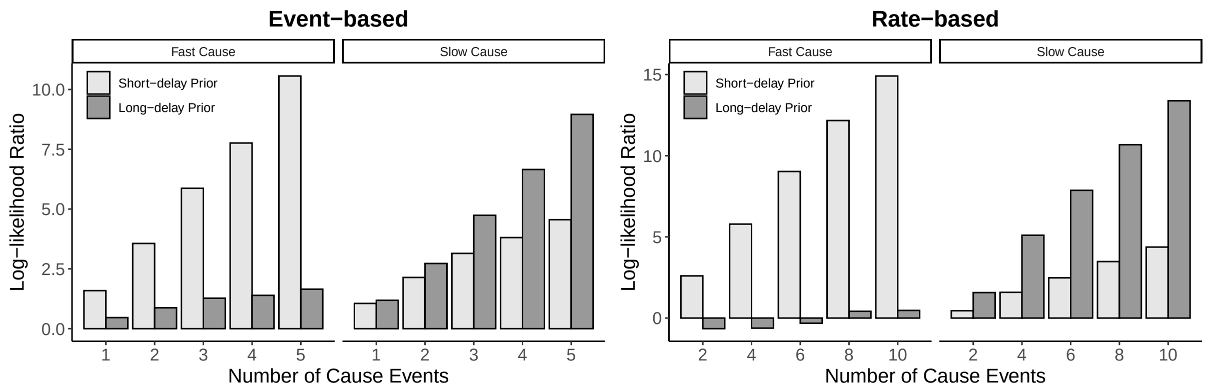


Figure 8

How log-likelihood ratio changes under different delay prior.

(i.e., structure induction). Additionally, the log-likelihood of S_1 itself decreases as the delay length increases. This indicates that when faced with multiple potential cause candidates (i.e., diagnostic causal reasoning), the learner should tend to attribute the effect to the cause with the shortest delay. Similar patterns are replicated under the rate-based scheme (Figure 7b). We show in Appendix B how the time-scale invariance property would appear instead if we adapt the baserate and delay variance proportionally to the delay length.

Delay variance

To investigate the predictable-delay preference, we simulate stimuli arranged in a grid with $m_u = \{10, 20, 30\}$, $i_u = \{1, 3, 5, 7, 9\}$, $w_c = 1$, $k_b = 3$. As shown in Figure 7c, the event-based model's preference for S_1 over S_0 diminishes as the range of delays increases. It reflects that causal attribution is stronger when the delays are reliable. Additionally, the log-likelihood of S_1 decreases as the delay range expands, which suggests that when faced with multiple potential cause candidates, the learner should tend to attribute the effect to the cause with the most consistent or reliable delays. Similar results are observed in the rate-based scheme (Figure 7d). In Appendix C, we show how our model replicates the finding by Greville and Buehner (2010) that the effect of variance persists as the learning duration increases.

Delay expectation

To investigate the influence of prior beliefs on causal judgments, we introduced two different delay prior conditions instead of using the above uniform delay priors. For the "short-delay prior", we set μ to be sampled from a Gamma distribution with a mean of 10 and a standard deviation of 1, resulting in an assumed delay expectation of 10$\pm$1. Conversely, for the "long-delay prior", we assume μ is sampled from a Gamma distribution with a mean of 20 and a standard deviation of 1, representing an assumed delay expectation of around 20$\pm$1. Other model parameterizations remain the same as before.

For the synthetic data, we constructed scenarios in which a slow cause always produced an effect with a delay sampled from $m_u = 20$, $i_u = 1$, while a fast cause always produced an effect with a delay sampled from $m_u = 10$, $i_u = 1$. We set $w_c = 1$ and $k_b = 3$ when simulating the data. Figure 8 demonstrates that in the fast-cause scenarios, the preference for S_1 in both event-based and rate-based models is stronger when one starts with a short-delay prior, while in the slow-cause scenarios, the preference for S_1 is stronger when one starts with a long-delay prior. This confirms that the model will learn causal relations more quickly when their time course aligns with expectations. It is worth noting that the models' tendency to favor the slow cause under the long-delay prior is not as strong as the tendency to favor the fast cause under the short-delay prior, highlighting the natural advantage of shorter delays.

Human Performance in Learning from Continuous-time Evidence

We now reanalyze seven previous datasets that contain human performance in a variety of temporal causal learning tasks, as shown in Table 1 (the corresponding hypothesis spaces of causal structures are summarized in Figure 9). We will demonstrate that our framework can accommodate all the variations across the scenarios probed across these tasks. Furthermore, we will demonstrate a robust alignment with participants' judgments and the predictions of the framework across these scenarios.

It is worth noting that all the datasets we model contain "interventions" — most of them provided as pre-set interventions, while Gong et al. (2023) allowed participants to intervene actively at moments of their choosing. The contingency-based learning literature emphasizes the differences between intervention and covariation-only data (Bramley et al., 2015; Lagnado & Sloman, 2002; Lagnado et al., 2007; Pearl, 2000; Sloman & Lagnado, 2005; Waldmann & Hagmayer, 2005). In the contingency setting, interventions act as "graph surgery" making observationally equivalent (aka. Markov equivalent structures) distinguishable. For example, casual structures such as $X \rightarrow Y \rightarrow Z$, $X \leftarrow Y \leftarrow Z$, and $X \leftarrow Y \rightarrow Z$ have the same co-variation patterns (X and Y are always correlated, as are Y and Z, and X and Z are unconditionally correlated but become uncorrelated once Y is controlled for), under different parameterizations all three models are equally good accounts of contingency data with these independencies. People are able to select interventions that allow them to distinguish these structures in causal learning experiments (Bramley et al., 2015; Coenen et al., 2015). However, event timing can break the Markov equivalence deadlock in some cases (Lagnado et al., 2007). For example, the temporal pattern $X - Y - Z$ is privileged under the first structure, $Z - Y - X$ in the second, and $Y - X - Z$ or $Y - Z - X$ in the third structure (even though high base rates or preventative connections may complicate the pattern). Delay information will ultimately tend to support whatever causal hypothesis most parsimoniously links the events, assigning them the highest joint likelihood by implying the shorter, more reliable and expected causal connections. However, whether the observationally most likely model is truly causally correct, i.e., that the statistical pattern is not due to some unobserved or latent variable (cf. Valentin et al., 2022), can be confirmed definitively through the use of interventions.

Continuous, effect specified

Lagnado and Speekenbrink (2010). Our first case study revisits the "earthquake" experiment conducted by Lagnado and Speekenbrink (2010). Participants were asked to investigate the effects of three types of seismic waves (red, yellow, and green) on the occurrence of earthquakes. Unbeknownst to them, only one of the three types of

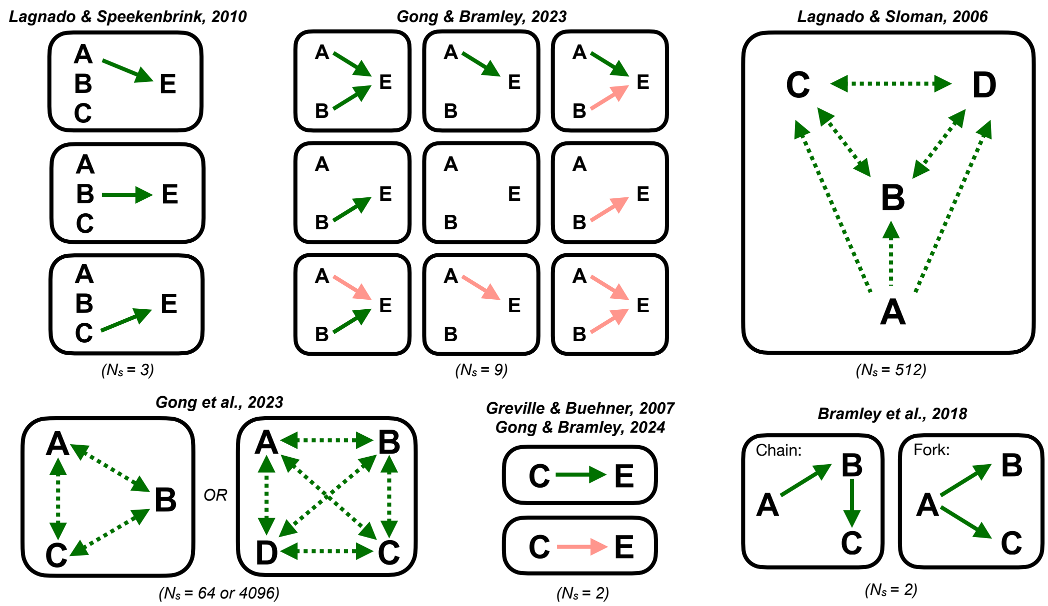


Figure 9

The hypothesis spaces used in different studies. Green arrows represent generative links and pink arrows represent preventative links. Dashed arrows $A \rightarrow B$ represent two possibilities between two variables A and B: unconnected or $A \rightarrow B$, and dashed bidirectional arrows represent four possibilities between two variables A and B: unconnected, $A \rightarrow B$, $B \rightarrow A$, or $A \leftrightarrow B$. Exogenous links (base rate) are ignored in all graphs.

waves (referred to as the cause) actually made earthquakes occur, while the other two types (referred to as lures) had no effect on the earthquakes. In each trial, the cause wave occurred 10 times and had a probability of 80% of resulting in an earthquake. Two other lure causes also occurred 10 times each but had no effect on the earthquake. Two factors were manipulated across trials within subjects: the delay length — the time between a cause event and its effect event — could be either short (3±0.1 s) or long (6±0.1 s); the probability of intervening events — how often the two other lure causes occurred between a real cause and its effect — could be either high (65%) or low (35%). Four additional earthquake events were sampled at random time points to serve as the base rate. The trials lasted for an average duration of 169±84 s for the short-delay condition and 318±157 s for the long-delay condition. Since the lures occur in the interval between the cause and effect, a learner might mistake the lure for the true cause if they do not pay enough attention to temporal information. Lagnado and Speekenbrink (2010) showed participants' ability to figure out the genuine cause: they assigned higher ratings to the genuine cause compared to the two lures. Meanwhile, the judgments were influenced by the probability of intervening events rather than the mere length of delay (see Figure 10a).

There was a mixed set of factors predicting whether a short-delay intuition should emerge. This study matched the base rate effect events (the long-delay condition had a lower base rate due to a fixed number of background effects over a longer observation period). However, the within-subject design provides a shared context liable to undermine the invariance in the minds of participants. Moreover, the long-cause and short-cause stimuli were created independently, rather than being directly scaled from one another. Notably, the standard deviations for the two conditions were identical (0.1 s), meaning the long-delay condition had a smaller relative variance than the short-delay condition, making the causal relationships easier to identify from a normative perspective. As such, a formal modeling procedure is necessary to determine the rationally predicted direction of the delay effect in this task.

Participants were asked to provide both “absolute” and “comparative” ratings for the causal properties of each wave. The “absolute” rating allowed participants to independently rate each wave, while the “comparative” rating required participants to allocate ratings for the three waves such that their ratings summed up to 100. Both types of ratings revealed the same pattern with only a main effect of the probability of intervening events but not the delay length (see Figure 10a). Here we model the “comparative” rating, which can be interpreted as a comparison of the posteriors associated with three causal structures as shown in Figure 9 (assuming that they each have an equal prior). None of the delay or other parameters were explicitly disclosed to the participants in the instructions, and the task included abstract visualizations and gamification features, which might suggest to participants that it was more like a game than a real-world situation. Therefore, we here do not speculate the prior knowledge participants may use. We use the prior for parameters as follows: $w_c \sim U(0, 1)$ for the cause probability (and $\lambda_1 \sim U(0, 1)$ for the rate-based model), $\mu_b \sim U(0, 100)$ for the base rate mean (and $\lambda_0 \sim 1/U(0, 100)$ for the rate-based model), $\mu \sim U(0, 100)$ for the cause delay (or influence) mean, and $\sigma^2 \sim U(0, \mu^2)$ for the cause delay (or influence) variance. We generated a simple Monte Carlo sample of size $m = 100,000$ to approximate the Bayesian inference process. We will use $m = 10,000$ for the remaining datasets (unless all parameters were assumed to be known to the model) because these datasets had either fewer unknown parameters or narrower parameter ranges compared to this study.[14]

As shown in Figure 10a, both event-based and rate-based schemes successfully identified the genuine cause in each condition, similar to the participants. Also consistent with participants' responses, the models were more influenced by the probability of intervening events than by delay length. As mentioned above, a mixture of factors in the study could result in different predictions as to whether there should be a short-delay or long-delay preference. As a result, the rational model predicted similar strength judgments in both conditions, with a slight tendency to favor the long delay (Figure 10a). This is driven by the relatively smaller variance in the long-delay condition. Although this tendency did not manifest clearly in participants' responses, we can see how helpful a rational model can be in isolating the influences of different factors, which is valuable for designing experiments and further developing process-level models to account for things like memory storage and perceptual noise. Further trial-level comparisons were not conducted with this dataset because of the lack of trial-level human judgments.

Gong and Bramley (2023). A somewhat similar dataset was collected in Gong and Bramley (2023). Participants were presented with a causal device consisting of one target component (Effect E) and two control components (Cause A and B). The

[14] The rate-based scheme requires a decision on the time bin configuration. The bins should be fine enough to satisfy that a cause should happen in a time bin before the time bins capturing most of its effects. Here we used 1 second for Gong and Bramley (2023), Gong et al. (2023), Lagnado and Sloman (2006) and Lagnado and Speekenbrink (2010) and 1 day for Gong and Bramley (2024) and Greville and Buehner (2007). Both choices can be regarded as natural. We used 300 milliseconds for Bramley, Gerstenberg, Mayrhofer et al. (2018) given that more coarse choices would compromise the accuracy.

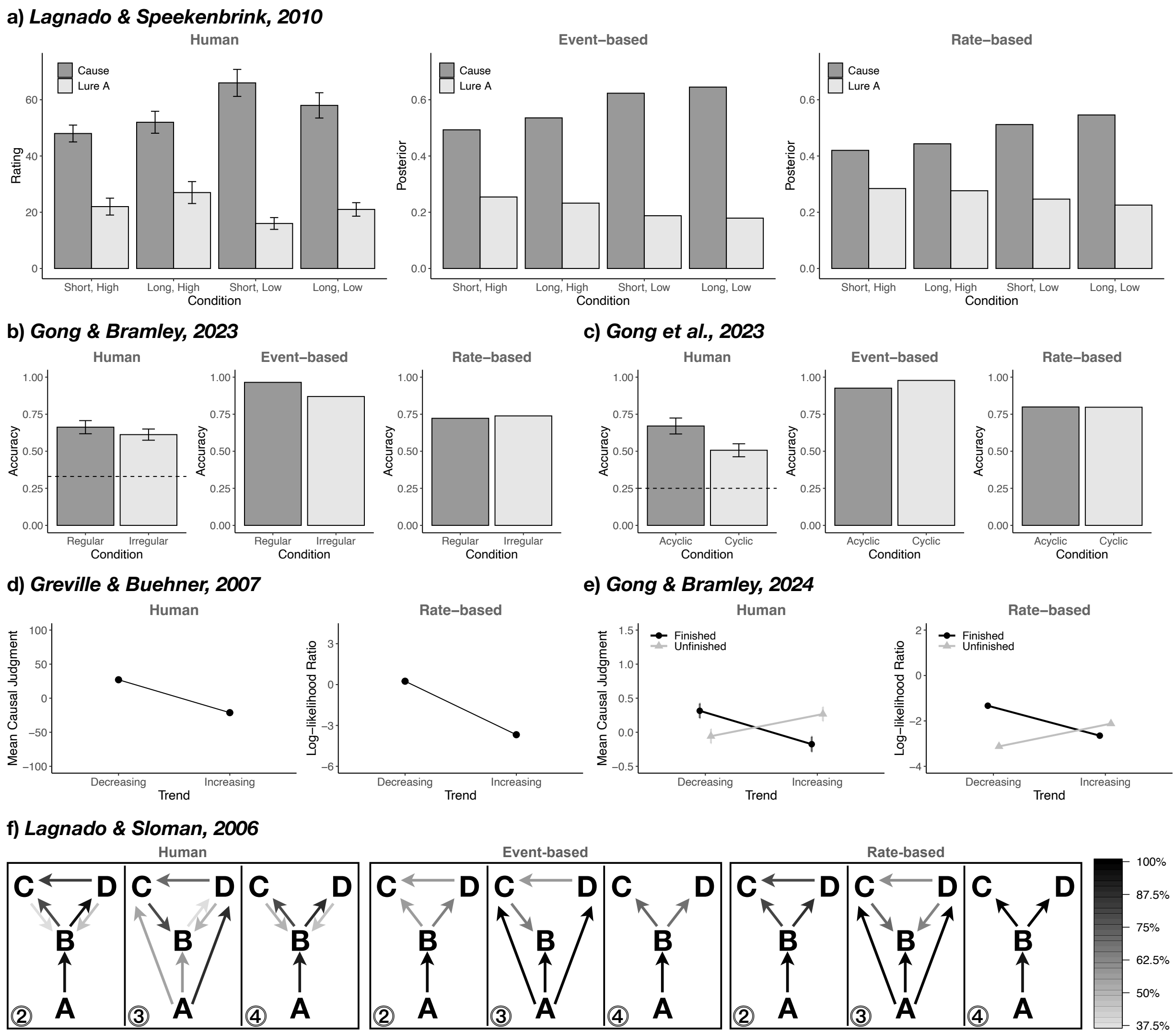


Figure 10

Aggregated results for six datasets. A temperature parameter of 15 was applied to Lagnado and Speekenbrink (2010) for visualization. Only one of the lures was reported in the original paper, presuming that the other lure could be calculated given the constraint in comparative ratings (LureB = 100 - Cause - LureA). Horizontal dashed lines in Gong and Bramley (2023) and Gong et al. (2023) indicate the chance-level performance. Ratings in Greville and Buehner (2007) are reversed so that they are aligned with Gong and Bramley (2024) where positive numbers indicated harmful influence and negative numbers indicated beneficial influence. The shading corresponds to the percentage of link selection in Lagnado and Sloman (2006); only links endorsed by more than 8 out of 24 participants or more than 50% chance by the models (after a softmax fitting) are shown; the occurrence orders were A-B-D-C, A-D-C-B, and A-B-CD (with the latter two occurring simultaneously) in Condition 2-4, respectively.

relationships between each control component and the target component could be generative, non-causal, or preventative, resulting in 9 possible causal structures (see Figure 9). A generative cause event would always produce an effect event after a gamma distributed delay of 1.5±0.5 s. A preventative cause event would cancel any upcoming effect events during a subsequent gamma distributed prevention window of 3±0.5 s. The effect component could also activate spontaneously. Participants were randomly assigned to the regular base rate or the irregular base rate condition. Each base rate event occurred semi-periodically, with gamma distributed 5±0.5 s delays after the previous one in the regular condition, or unpredictably with exponentially distributed 5±5 s in the irregular condition. Participants watched the device being intervened on by an artificial agent for a total duration of 20 s during which there would always be three interventions on A and three on B.

In the experiment we model here, participants were given video training to experience the delay parameters mentioned above. They were also explicitly told about the mean of the delays in the instruction. Therefore, we make the matching assumption that the model is also aware of these parameters. Specifically, for generative causes, we set $w_c = 1$ (and $\lambda_1 = 1$ for the rate-based model), the generative delay $\mu = 1.5$ and $\sigma^2 = 0.25$. Regarding the base rate, we assume a mean of $\mu_b = 5$ (and $\lambda_0 = 1/5$ for the rate-based model).

In the case of preventative causes, the event-based scheme assumes the duration of preventative windows follow a gamma distribution with a mean of 3 and a variance of 0.25 (Figure 11a). All events occurring within a prevention window are assumed to be canceled. The rate-based scheme models the dynamics of preventative influence. It should be noted that the actual preventative mechanism employed here does not involve an incubation or decay process. Rather, the preventative window stays at its maximum level, effectively canceling all effects, for a certain duration. As such, under the rate-based scheme we accommodate this mechanism by modeling the preventative causal influence using the gamma cumulative density function, as illustrated in Figure 11b, and assuming a maximum level denoted as $\xi = 1$.

The task instructions in Gong and Bramley (2023) imply three mechanistic features of the scenario that can be implemented in the event-based scheme but less easily under the rate-based. Firstly, a single generative cause event only produced one additional occurrence of the effect component, which is consistent with the setup of the earthquake experiment (Lagnado & Speekenbrink, 2010) described earlier. Secondly, in the regular condition, the base rate events occurred semi-periodically. Therefore, instead of utilizing a memoryless exponential distribution, the event-based model can employ a gamma distribution (with a mean of 5 and a variance of 0.25), to model the semi-predictable delay between consecutive base rate events. In contrast, since the rate-based model does not differentiate between effects generated by base rate events or generative causes, it is unable to leverage the regularity of the base rate and thus treats the regular and irregular conditions in the same manner. The third rule pertains to the preventative window. In the generative process, it was the case that within a fixed preventative window, all expected effects would be canceled, while any expected effects after the window would remain unaffected. Consequently, the size of the true preventative window has to be smaller than the interval between a preventative cause event and its nearest subsequent effect event E'. The absence of an effect expected to occur after this E' can no longer be attributed to the preventative causal influence. On the other hand, the rate-based model represents prevention as a probabilistic influence, defining a soft window rather than a strict, deterministic one.

As shown in Figure 10b, aggregately the event-based scheme has higher accuracy compared to the rate-based scheme in identifying the causal models in Gong and Bramley (2023). It also shows a similar pattern of performing better in the regular condition compared to the irregular condition, which aligns with human performance. In contrast, the rate-based scheme demonstrates a slight tendency to perform better in the irregular condition, potentially attributable to the alignment of the base rate mechanism with the model's assumptions.

For this and the subsequent datasets, we measured two types of correlation between model judgments (the posterior probability of each answer) and human judgments (the proportion of participants providing each answer). The first type is the Pearson correlation, which incorporates a softmax parameter to account for the stochastic nature of judgments (Luce, 1959).[15] We used a single parameter that we fit across all conditions for each dataset. The second type is the Spearman correlation, which assesses the ranking agreement between human and model judgments. This provides insight into how well the model captures the human dataset without introducing an additional free parameter. The results are depicted in Figure 12a and 13a. Both the event-based and rate-based schemes successfully captured human judgments, regardless of whether the conditions were regular or irregular. The event-based model demonstrated slightly superior correlations compared to the rate-based schemes, suggesting that participants may have taken into account at least some of the particular mechanistic constraints discussed above, during their reasoning process.

We demonstrate here that people's judgments reflect rational considerations. However, this does not mean the rational equations provide a good process-level account of how people make these judgments. It is plausible that people relied on other algorithms to approximate the rational solution. Gong and Bramley (2023) explored the types of approximation algorithms people may use to choose among nine causal hypotheses after observing a more-than-countable number of events short observation period. They found that a summary-statistics approach, based on structurally local computation using temporally local evidence, provided a better fit to participants' judgments than the

[15] The softmax parameter θ was used to maximize the log-likelihood between models' and participants' choices in Gong and Bramley (2023), Gong et al. (2023) and Lagnado and Sloman (2006), where participants' answers were binary about whether each causal connection existed or not. The parameter was used to maximize the linear correlation based on a non-linear transformation $y = sign(x)|x|^\theta$ in Bramley, Gerstenberg, Mayrhofer et al. (2018), Gong and Bramley (2024) and Greville and Buehner (2007) where participants provided continuous ratings for how likely each connection existed or how strong each causal strength was.

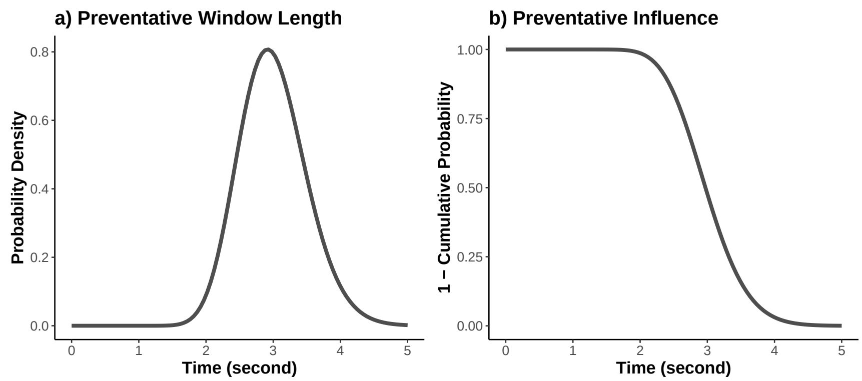


Figure 11

The preventative windows and preventative influences used to model Gong and Bramley (2023). a) The event-based scheme assumes the length of each preventative window is sampled from a gamma density function. b) To approximate this within the assumptions of the rate-based scheme, we assume the probability of prevention (what proportion of the effects will be prevented) is an inverse cumulative gamma function, reflecting a decreasing preventative influence over time.

normative model. Readers interested in further details are referred to Gong and Bramley (2023). We further discuss process-level models as the future steps in the General Discussion.

Continuous, effect unspecified

Gong et al. (2023). When the effect variables are left unspecified, the number of potential causal structures increases quickly in the number of relata. Even when considering only generative relationships, there are four possible relationships between two variables: one-directional, reverse one-directional, bidirectional, and unconnected. Consequently, for three variables, there are 64 potential structures, and for four variables, there are 4096 potential structures (refer to Figure 9). Gong et al. (2023) investigated how individuals learn about causal structures drawn from the large 3- and 4-variable hypothesis spaces by actively intervening in a causal system.

Although this is an active learning task, we use this model to analyze participants' judgment patterns rather than their intervention patterns. The rational framework we present here is an inference model, so it makes no distinction between learning from one's own interventions and learning from another person's interventions (these distinctions

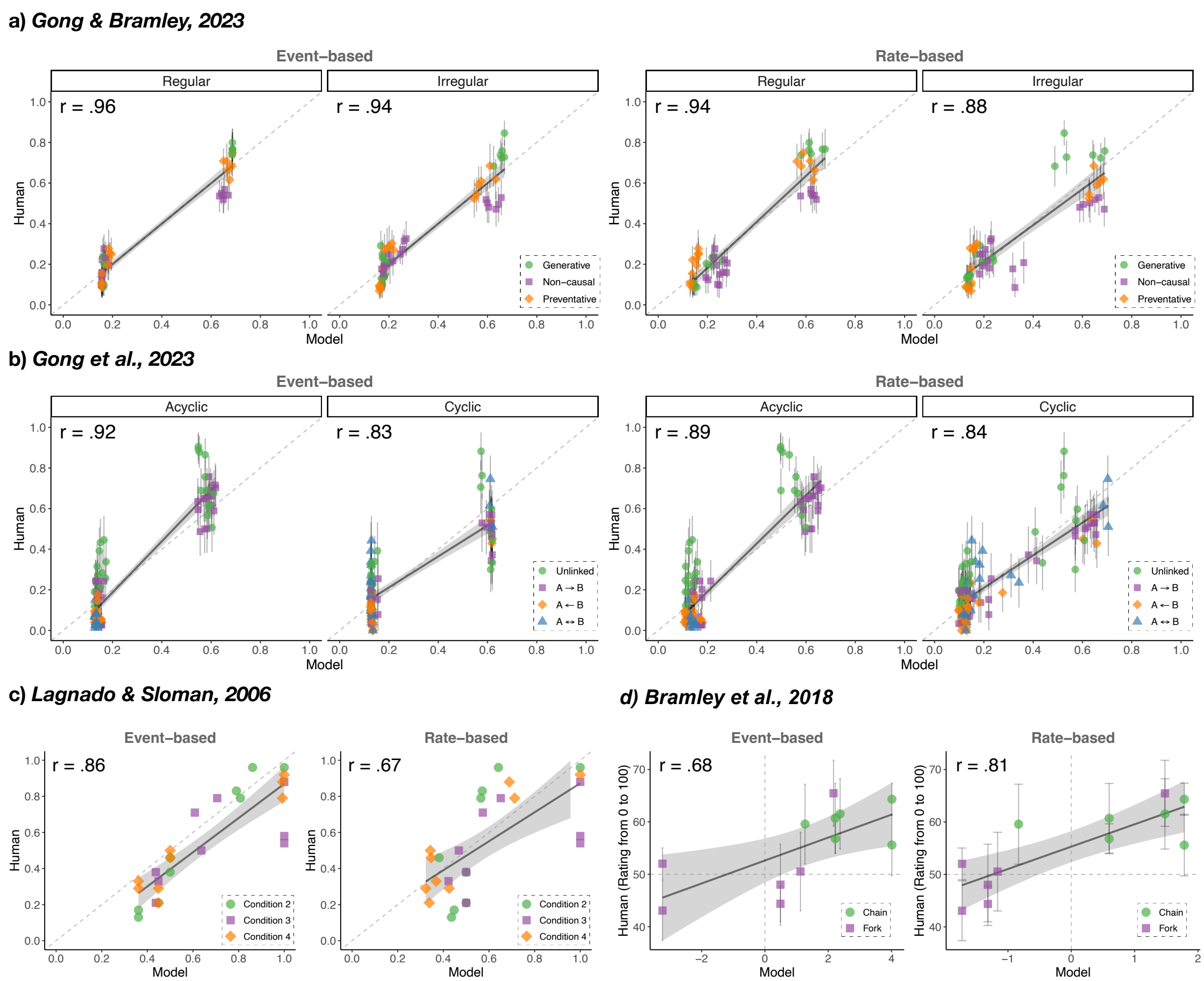


Figure 12

Pearson correlations between model and human judgments. Y-axes indicate the proportion of human judgments in Gong and Bramley (2023), Gong et al. (2023), and Lagnado and Sloman (2006). Error bars indicate 95% confidence intervals of human judgments in the dataset whenever the raw data are available. Each individual point is the averaged response for a particular stimulus.

could be probed with a process-level model; Markant & Gureckis, 2014). To further develop a rational account for how a learner should intervene to learn more effectively and efficiently, we need to combine the current inference model with preposterior calculations using information gain measures from information theory. Although this problem is beyond the scope of this paper, readers are further referred to Gong et al. (2023) which makes and tests proposals as to if and how people do this in continuous-time contexts.

In the Gong et al. (2023) experiment, half of the ground truth structures contained a cycle and half did not. Delay regularity was manipulated between participants. In each causal system, for causally related components, an activated component would probabilistically trigger the activation of each of its effect components once after a delay of 1.5±0.1 s in the regular condition, or after a delay of 1.5±0.7 s in the irregular condition. All causal connections worked 90% of the time, and none of the components activated spontaneously (i.e., there were no base rate activations). Participants were provided with six opportunities to activate a component in the system during a 45-second interval. Considering the possible numerous connections and the cyclic structures, the number of events recorded in this dataset was also significantly higher compared to the aforementioned Gong and Bramley (2023).

Participants in the study were pre-trained about the causal parameters with a similar procedure as in Gong and Bramley (2023), and we hence assume that models also know these parameters: $w_c = 0.9$ (and $\lambda_1 = 0.9$ for the rate-based model), $\mu = 1.5$, and $\sigma^2 = 0.01$ or $\sigma^2 = 0.49$ depending on the specific regular or irregular condition. No base rate was assumed by either model.

Human results showed a main effect of the structure cyclicity but no main effect of the delay regularity, probably due to that the difference between regular and irregular settings was not pronounced enough (Gong et al., 2023). Therefore, we here focus on the results based on the cyclicity factor alone. In contrast to humans who performed better in the acyclic condition than the cyclic condition, the event-based model demonstrates better performance in the cyclic condition compared to the acyclic condition (Figure 10c). This reflects the fact that the event-based model is able to leverage the larger amount of event information available in the cyclic structures via the one-effect per cause mechanistic constraint (Gong et al., 2023). Conversely, the rate-based model does not demonstrate the same tendency. Due to not enumerating actual causation pathways, this model fails to leverage the information available in abundance of cyclic events as effectively as the event-based model does.

In terms of both correlation measurements, the event-based model demonstrates better performance in capturing human judgments in acyclic structures, while the rate-based model performs better in capturing human judgments in cyclic structures (see Figure 12b and 13b). This may suggest that as the number of events increases, the exact enumeration computations become infeasible for people, necessitating the relaxation of the one-cause–one-effect constraint within the event scheme to enable more efficient approximations. The rate-based model's capability to capture human cyclic judgments highlights its ability to deal gracefully with larger number of events, providing a not perfect but more efficient approach in such scenarios.

Once again, the analysis here demonstrated participants' rational thinking in solving the online causal structure learning problem, but this does not imply that participants can perform exactly as the rational models do. As shown in Figure 10c, participants' accuracy was lower than that of the models, and their accuracy decreased with the increasing number of events (Gong et al., 2023). This suggests that they may not be able to store and utilize the temporal information of all observed events as effectively as the rational framework.

Episodic, effect specified

Greville and Buehner (2007). We refer to "episodic evidence" as evidence gathered from multiple independent causal systems of the same type, where each sample has its own timeline. Episodic evidence can be seen as a combination of contingency and temporal information (Greville & Buehner, 2007). It involves the observation of multiple independent individuals, but with each observation lasting over a specific time period (Figure 1b). Research on episodic evidence often focuses on cases in which each type of event occurs at most once within the observed period (Bramley, Gerstenberg, Mayrhofer et al., 2018; Gong & Bramley, 2024; Greville & Buehner, 2007; Lagnado & Sloman, 2006). This means that the evidence within each individual's experience may not be very informative. However, by considering multiple cases, the reasoner can compensate for the

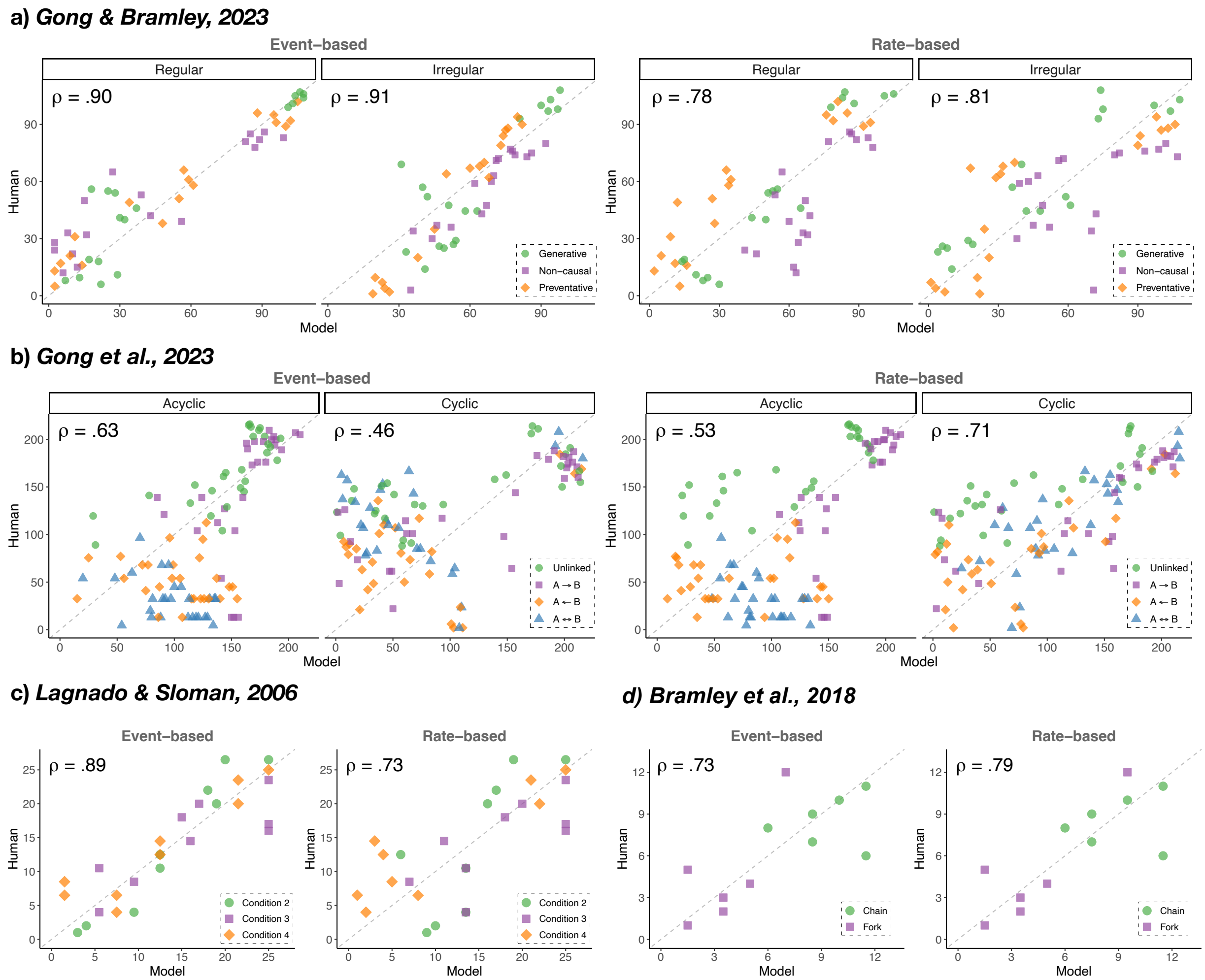


Figure 13

Spearman correlations between model and human judgments. Axes are the ranks of judgments.

limited information within each instance and make more informed conclusions about their common causal structure.

In Greville and Buehner (2007), participants were asked to examine the influence of a ray treatment on the survival of bacterial cultures. Bacterial cultures were randomly assigned to the experimental group, which received a ray treatment at Day 0, or the control group, which did not receive any treatment. Each group consisted of 40 samples. Bacterial cultures were observed from Day 1 to Day 5. The number of new deaths occurring each day was recorded.[16] Participants were asked to rate whether they perceived

[16] The data were displayed as tabular records indicating whether each culture was still alive or not in

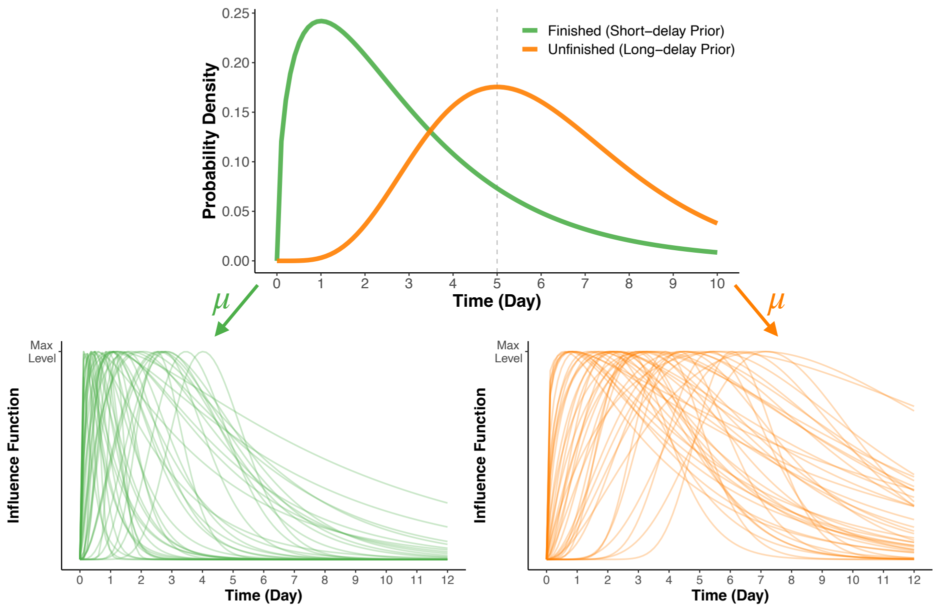


Figure 14

Short-delay and long-delay priors regarding the timing of when the cause will take effect on average (Gong & Bramley, 2024; Greville & Buehner, 2007). The parameter μ is sampled from different prior distributions to form different causal influence functions.

the treatment as harmful or beneficial based on the observed outcomes in both the experimental and control conditions. The control condition always demonstrated a relatively constant death rate over time (e.g., 8, 8, 8, 8, 8), while the daily death rate in the experimental condition was manipulated to exhibit either a decreasing or an increasing trend. Results showed, after controlling for the total number of deaths over the 5-day period, participants judged the treatment as more harmful if there were more deaths right after the treatment (a decreasing trend; e.g., 16, 12, 8, 4, 0), and more beneficial if there were more deaths towards the end of the observation period (an increasing trend; e.g., 0, 4, 8, 12, 16). As such, this study demonstrated that people not only care about the overall contingency data summarized from the entire observation period but also the detailed temporal dynamics (at a day-to-day level here).

Greville and Buehner (2007), and as summarized counts of how many cultures died each day in the later Gong and Bramley (2024). Given that the results of Gong and Bramley (2024) replicated the same "finished" condition as in Greville and Buehner (2007), we do not discuss the potential influence of formats here.

Gong and Bramley (2024). While agreeing on the impact of temporal dynamics on judgments, Gong and Bramley (2024) proposed that some settings could produce a different pattern than the traditional notion of contiguity (Greville & Buehner, 2007). In this task, if in some conditions learners tended to assume that the causal process may be ongoing, an increasing trend might signal that the treatment will ultimately prove harmful. Gong and Bramley (2024) presented participants with more such ambiguous data, where a majority of the forty samples were still alive on Day 5 (e,g,. 0, 1, 1, 3, 5 in the experimental condition and 1, 3, 2, 2, 2 in the control condition). Participants in the "Unfinished" condition were informed that the observation had not yet concluded, while participants in the "Finished" condition were told that the observation had finished (as in Greville & Buehner, 2007). Results in the Finished condition replicated Greville and Buehner (2007). However, in the Unfinished condition, participants interpreted an increasing trend in deaths as indicative of harm caused by the treatment, and a decreasing trend as indicative of benefit (see Figure 10e).

These findings highlight the influence of instructional cues on participants' inductive biases and how they interpret the observed data. To model the human judgments, we here assume that instructions tend to influence the learner's prior expectation about causal delays as well the use of data. As shown in Figure 14, if participants are informed that the experiment ends at Day 5, they may tend to form a prior belief that the relevant causal influences are expected to occur within 5 days. When participants were led to believe that the observation had not finished, they anticipated the possibility of longer causal delays. Here we assume that, for the Finished instruction (Gong & Bramley, 2024; Greville & Buehner, 2007), the causal delay (or the expected time of the influential function in the rate-based context) μ is sampled from a gamma distribution with mean of 3 (days) and a variance of 6. For the Unfinished instruction (Gong & Bramley, 2024), μ is sampled from a gamma distribution with mean of 6 (days) and a variance of 6. While a range of values might be reasonable, we chose these because range of 0 to 5 would cover most of the sampled μ under the Finished instruction (83%) while covering only a minority of the sampled μ under the Unfinished instruction (38%, Figure 14).

The observed death of each bacteria culture in Greville and Buehner (2007) and Gong and Bramley (2024) could result from either the treatment or natural death (i.e., the base rate). Given that the cultures which died out in the same day were not distinguishable from each other, we here focus on the rate-based scheme and consolidate the data as shown in Figure 4e. Since the data were collapsed, the rate of how many events happened per day depends on the total sample size (i.e., forty in both studies). We assume that participants selected between the "harmful" and "beneficial" options by comparing the likelihood between a generative structure and a preventative structure (see Figure 9). To model the data, we set $\lambda_0 \sim U(0, 40)$, $\lambda_1 \sim U(0, 40)$ since there were at most forty cases in each group. We set $\xi \sim U(0, 1)$ for the max level of preventative influence (i.e., the beneficial influence). Similar to previous datasets, we set the variance $\sigma^2 \sim U(0, \mu^2)$.

The aggregated results are shown in Figure 10d and 10e. Participants' inclinations in Greville and Buehner (2007) and Gong and Bramley (2024) were captured by the model. Under the Finished instruction (Gong & Bramley, 2024; Greville & Buehner, 2007), participants and the model both treated decreasing trends as more harmful than increasing trends, lining up with the contiguity-driven explanation. Under the Unfinished instruction (Gong & Bramley, 2024), participants and the model both treated increasing trends as more harmful than decreasing trends, lining up with the sensitivity to trends. Nevertheless, the model was less likely to demonstrate a cause as absolutely harmful (i.e., giving a rating above 0). This could be due to the fact that Gong and Bramley (2024) handcrafted the stimuli, where only a small number of bacterial cultures out of the sample (n=40) died during the observation, and the overall death rates were similar between the experimental and control groups (e.g., 3, 4, 2, 1, 0 in the experimental group and 2, 2, 1, 2, 3 in the control group). An ideal learner would expect the experimental condition to still exhibit a higher death rate towards the end of the observation if they were to claim a cause as an

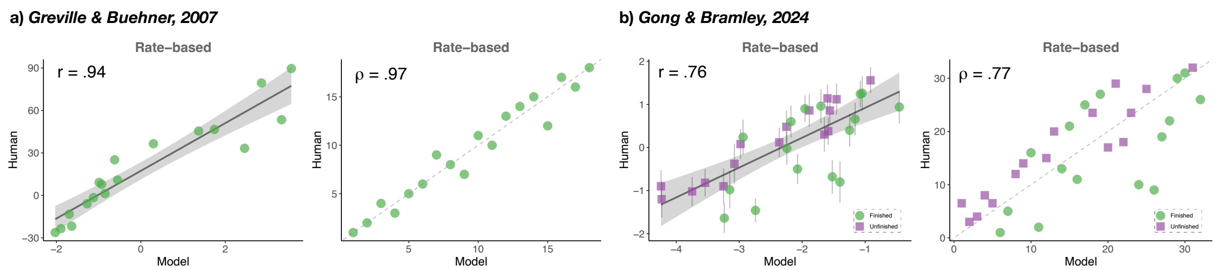


Figure 15

Pearson (r) and Spearman (ρ) correlations between model and human jugdments in Greville and Buehner (2007) and Gong and Bramley (2024). Error bars indicate 95% confidence intervals of human judgments in Gong and Bramley (2024).

absolute harm. Participants' deviations from this prediction could have several explanations. One speculation is that they did not simply represent harm as generative (causative of death) and benefit as preventative, like the model. They may assess whether a treatment is good or bad by also considering whether it advances or delays death, which is a sensible interpretation, but falls outside the scope of our current model. We will return to this point as a future direction in the General Discussion.

The trial-level results are shown in Figure 15. The rate-based model achieved a good fit with human judgments at the trial level in Greville and Buehner (2007) and Gong and Bramley (2024).

Episodic, effect unspecified

The final category we consider here is episodic evidence where the effect variable is unspecified (i.e., accepting a wider range of hypothetical structures rather than focusing on specifying one variable as the effect and finding its causes; Figure 9). The two datasets we consider here both involve scenarios where each kind of event can only happen once in each episode.

Lagnado and Sloman (2006). In Lagnado and Sloman (2006), participants were asked to imagine a situation in which a computer virus can spread through a network and told that the time at which a computer revealed its infection could occur after a variable delay, so later than the time at which the computer became infected. Participants were told that each connection, if it existed, would spread the virus 80% of the time, and the virus could not reach a computer unless it had been sent from another (e.g., no spontaneous base rate infections would occur). Participants watched 100 clips, each showing an event sequence in which the virus appeared in different computers, and were asked to judge the existence of various potential causal links (i.e., directed network connections) in the system (see Figure 9).

The experiment included four conditions (in Condition 1 all events happened simultaneously so won't be modeled here), but the underlying ground truth structure was consistently: A was the cause of B, and B was the common cause of C and D. This meant that in each condition, computers C or D would never become infected without computer B being infected. Since the actual infection time was varied and unknown, the presumed rational solution is to rely on the conditional probability. However, the timing of virus appearance in each computer could be misleading. For example, in Condition 3 where 50% of trials followed the order of $A - D - C - B$, participants tended to judge the links $A \rightarrow D$, $D \rightarrow C$, and $C \rightarrow B$ were more likely to exist than other links (see Figure 10f). Their answers cannot explain the other 50% of trials when only AB, ABC or ABD happened. This suggests that people's reliance on temporal information is so strong that it could not, in this case, be overshadowed by the contingency information. As such, in order to capture human judgments, what we will model here is not the objective rational solution, but the rational solution having disregarded the instruction that the temporal information is irrelevant.[17]

There was a one-second delay between events in subsequent time steps (t_1, t_2, t_3, t_4; see Table 2 in Lagnado & Sloman, 2006). As such, each trial lasted 4 s. We model the dataset using the parameters $w_c = 0.8$ (and $\lambda_1 = 0.8$ for the rate-based model) as in the instruction, $\mu \sim U(0, 10)$, $\sigma^2 \sim U(0, \mu^2)$. The base rate is assumed to be zero. In this

[17] Note that after personal contact, we learned that the condition-by-condition human judgments published in Table 3 of Lagnado and Sloman (2006) were incorrect. Readers may refer to p.188 in Pacer (2016) or the GitHub repository of this paper for the corrected table.

dataset, the main difference between event-based and rate-based schemes is that the former address the rule that an event can only occur once for a specific equipment, which is consistent with the experimental design. Consequently, the event-based scheme outperforms the rate-based scheme in accurately capturing human judgments, as shown in Figure 12c and Figure 13c. Both models capture the aggregated human judgment patterns, as shown in Figure 10f. This includes the Condition 3 mentioned above, despite the fact that the models also tend to add $A \rightarrow B$ and $A \rightarrow C$ links which can help rationally account for the trials when B and C happened without D. Our framework demonstrates an ability to explain the phenomenon that temporal information can outweigh contingency information in human causal judgments.

Bramley, Gerstenberg, Mayrhofer et al. (2018). Bramley, Gerstenberg, Mayrhofer et al. (2018) tested whether people can differentiate between two causal structures, chain and fork (see Figure 9), solely using delay information. Each trial consisted of 12 episodes, wherein events always occurred in the order $A - B - C$. However, there were variations in the delay variances between structures. In the chain structure ($A \rightarrow B \rightarrow C$), the delay variance between B and C was small, whereas the variance between A and C was large, as it encompassed the variability across both causal connections. Conversely, in the fork structure ($B \leftarrow A \rightarrow C$), the delay variance between A and C was small, while the variance between B and C was large, as there was no direct causal link between the two variables. Participants were asked to judge by distributing 100 percentage points across the two structures.

In contrast to previous datasets, we used the "independent delay" parameterization, as described in the original paper (Bramley, Gerstenberg, Mayrhofer et al., 2018), which allowed for there to be distinct delay distributions between different links in the causal structure. This means that to choose between chain and fork, we only need to model the delays between B and C in the chain hypothesis and the delays between A and C in the fork hypothesis (since both hypotheses share the $A \rightarrow B$ connection). Each episode lasted

3 s. We assume $w_c = 1$ (and $\lambda_1 = 1$ for the rate-based model), $\mu \sim U(0, 10)$, $\sigma^2 \sim U(0, \mu^2)$, and no base rate.

Results are shown in Figure 12d and 13d. Both models captured the general pattern of human judgments. The rate-based model demonstrated a better fit to human judgment compared to the event-based model. The event-based model showed a overall bias towards chains (judging all chain devices as chains and also judging some forks structures as chains). This is as expected and is due to the fact that $A - C$ delays (calculated under the fork hypothesis) were always longer than the $B - C$ delays (calculated under the chain hypothesis) in the stimuli. As a result, the ceteris paribus preference for the chain structure can be interpreted as an example of favoring the relatively shorter delay. In contrast, from the same evidence, the rate-based scheme, which assumes macro causal dynamic changes, has a greater tolerance for delays of causal influence than the event-based scheme. Participants may also have reasoned pragmatically that around half of the ground truths would be fork structures and so shifted their response threshold to favor the fork response more than the evidence supported.

We note here that the computational cost of the rate-based model is not always lower than that of the event-based model. In the case of this dataset, the rate-based model may actually be more computationally demanding depending on the granularity of the time bins it uses. Conversely, the event-based model benefits from the fact that each type of event occurs only once in each episode, resulting in a small number of causal pathways to consider within each hypothesis.

General Discussion

In this paper, we developed a rational framework for the use of temporal information in causal inference. The framework leverages stochastic processes from the Poisson-Gamma family to model the (in)dependencies between events in time and drive selection and parametrization of causal structure hypotheses. To achieve this, we extended the causal graphical model formalism to incorporate likelihood functions for temporal dynamics, before inverting these likelihoods to derive structural conclusions from evidence. We show how this general approach uncovers the underlying causal structure in all manner of complex continuous-time datasets. The framework is applicable to a wide range of temporal causal learning, associative and operant learning tasks, including scenarios where evidence comes in a long continuous timeline or from many shorter independent timelines, or when the causes can produce either one specific effect event or multiple effect events. As we demonstrate in modeling existing datasets, the framework accommodates variations such as the size of the hypothesis space, the involvement of background activity, preventative connections, cyclic dynamics, and whatever other information learners have about relevant causal delay distributions. To our knowledge, this is the first general computational framework for learning causal structure from events unfolding in continuous time.

The framework anticipates three intuitions that have been frequently observed in human learning: causal attributions are, in general, stronger to the extent that the delays between a putative cause and effect tend to be shorter, more consistent, and more in line with preexisting mechanistic expectations. We demonstrate the rationale behind these intuitions falls directly out of the Bayesian framework, explaining why it makes sense for them to coexist and why a preference may fail to manifest itself under certain circumstances. Additionally, we demonstrate that the framework helps explain behavioral patterns across a range of learning tasks from the last 20 years. We find a high degree of consistency between judgments derived from the rational framework and the aggregate behavior of human participants. These analyses suggest that people are not only capable of utilizing temporal information in diverse causal learning situations, but also that they do so in systematic, predictable, and approximately rational ways. By providing a unified computational framework, we are finally able to consolidate empirical studies spanning many different tasks and better clarify these tasks' relationship with widely studied associative and reinforcement settings. This model offers a design space for locating different tasks within the temporal causal learning field and provides guidelines for further investigation of causal cognition (Almaatouq et al., 2022).

In the remainder of the paper, we compare our model with previous temporal associative learning models, discuss why we think the pluralistic "dual-aspect" view we present here makes sense for describing human temporal causal learning, consider the relationship between continuous and discrete value representations, and lay out several future research directions.

A comparison to temporal associative learning models

We are not the first to point out the limitations of framing learning around trial-based contingencies, and consider how inferences operate on continuous-time data directly. This issue has been discussed in detail in the associative and animal learning literature. In particular, Gallistel and colleagues (Gallistel, 2021; Gallistel et al., 2014, 2019; Gallistel & Gibbon, 2000; Gallistel & Wilkes, 2016) argue that the way in which time is segmented into trials for analysis, as well as the duration considered as a single trial, can dramatically alter the predictions from contingency-based associative learning models; if researchers' choices depart from whatever intuitive structure and discretizations their subjects make in understanding their tasks, their models are doomed to lack the representational expressivity needed to capture their subjects' learning processes. Like us, Gallistel et al. propose that learning depends on the rate at which the effect occurs after a cause, or operant behavior, takes place (Gallistel, 2021; Gallistel & Gibbon, 2000; Rescorla, 1968). They further point out that temporal associative learning is not just determined by the frequency of temporal pairings (i.e., how often a presumed effect follows its presumed cause); it must also be sensitive to how often the effect occurs spontaneously without the cause's occurrence.

In their detailed experimental analyses, Gallistel et al. focus at the process-level on the pairwise attribution problem. For instance, Gallistel and Gibbon (2000) show that a response behavior is triggered when the ratio of rates with or without a stimulus exceeds a certain threshold. Gershman (2024) later suggest that a Rescorla-Wagner framework can be used to update the weights of different associative causes, by shifting from predictions about the presence or absence of effects to predictions about fluctuations in the rate of effects. Gallistel et al. (2014) also demonstrate that association strength depends on how much additional predictive *information* a presumed cause provides about its presumed effect, having accounted for the effect's base rate (Gallistel et al., 2014, 2019; Gallistel & Wilkes, 2016). In this more recent treatment, the casual inference no longer relies on detecting a rate change but instead on contrasting the random-timepoint → next-reward delay distribution against a cue → next-reward delay distribution, and using the entropy reduction between these distributions as a causal index.

There are obvious connections between Gallistel et al.'s theoretical ideas and ours: both embrace rate representations and contrast causal against baseline effect patterns. Their models, like our Bayesian approach, predict the phenomenon of time-scale invariance, because the associative strength depends only on the *relative*, not absolute rates or delays in situations with or without the cause (Gallistel et al., 2019; Gershman, 2024; Rescorla, 1968). However, apart from time-scale invariance, it is unclear whether an associative model can explain why learners favor causal explanations that posit causal delay durations which are relatively shorter, more in line with prior expectations, and more regularly timed.

One feature that differed across the experimental datasets was whether the effect was pre-specified. When the effect is specified, the task is to identify the true (positive or negative) causes of this effect (Gong & Bramley, 2023, 2024; Greville & Buehner, 2007; Lagnado & Speekenbrink, 2010). This is similar to the credit assignment problem in associative (or reinforcement) learning, where learners credit conditional stimuli (or interventions) for a particular unconditional stimulus (or a kind of reward). Temporal associative learning models can potentially provide predictions in these tasks. The advantage of the Bayesian framework is that it applies equally to the open-ended "structure learning" tasks, prototypical in the causal cognition literature, where nothing is a priori specified as a cause or an effect. These scenarios require more global reasoning (as well as interventional data) to solve reliably and the Bayesian framework helps clarify the circumstances where locally focused heuristics are or are not sufficient to arrive at the right global model (Bramley, Dayan et al., 2017; Fernbach & Sloman, 2009).

Another characteristic of the Bayesian account is its flexibility to deal with varied temporal dynamics. What it actually compares here is how well different causal explanations fit. This can include explanations making different assumptions about functional form as well as about structure. For instance, current temporal associative learning models implicitly assume that if a cause produces multiple effects, it will produce them at a constant rate during an effective time window with hard onset and offset boundaries (Gallistel et al., 2019; Gallistel & Gibbon, 2000; Gallistel & Wilkes, 2016). In contrast, we illustrate in this paper how the Bayesian approach can handle whatever hypotheses are articulated. For example, we modeled a case where changes in the effect's rate followed a latent, peak, and decay process continuously, but could contrast this with a uniform generation window or any other mechanistic hypothesis. Through the event-based scheme, it also allows for the incorporation of other mechanistic constraints, such as the case where a cause can generate only one effect, or the possibility that the where baseline effects are not unpredictable, but periodic. In such a situation where the base rate itself is a moving target, it is unclear whether a simple entropy reduction index (Gallistel et al., 2014, 2019) would provide a generalizable index of the power or strength of causal or relationship (cf. Cheng, 1997).

Note that all advantages we mention pertain to the flexibility of the model space that Bayesian inference is defined over. This is wholly compatible with the idea that at the process level, we rely on the mechanisms of pairwise association or reinforcement among other pragmatic, resource sensitive heuristics and approximations. Nevertheless, we hope this rational analysis is useful for mapping out the space of continuous-time learning problems including those classically used in associative learning tasks.

A pluralistic view

We presented two schemes, event-based and rate-based, in parallel throughout this paper but introduced both as manifestations of a broader Poisson-Gamma framework for conceptualizing interevent dynamics. The existence of a pluralistic view is not a new concept in the field of causal cognition. For instance, in research on token-level causal attribution, where individuals are asked to make judgments regarding what was responsible for particular event rather than causative of a class of events in general (Halpern, 2016), researchers have debated the relative importance of covariation versus process (Gerstenberg et al., 2021; Lombrozo, 2010; Sloman, 2005; Wolff, 2007). The question arises whether people prioritize imagining how the outcome would have changed if the cause had been different (Icard et al., 2017; Sloman, 2005), or if they focus more on determining if there was a genuine physical exchange between the cause and effect (Talmy, 1988; Wolff, 2007). Instead of relying on a single level of abstraction, people are pluralist, considering both the *occurrence* of the outcome and the *manner* in which it occurred (Gerstenberg et al., 2021). This paper focuses on a type-level causal learning rather than token-level causal attribution, meaning we can benchmark the quality of a judgment against the true causal generative model that they are learning about. We next give two reasons why a pluralistic perspective is also important in the domain of type-level causation, especially when it comes to temporal evidence based learning.

Mechanistic concerns

Causal structure learning can be driven by different types of evidence at different levels of abstraction. As we orient away from highly abstracted atemporal contingencies toward “raw” spatiotemporal dynamics, the richness of the data increases. Atemporal evidence discards a lot of information by discretizing into finite sets of categories and time points(Allan, 1980; Cheng, 1997; Griffiths & Tenenbaum, 2005; Perales & Shanks, 2007). For instance, researchers examined causal inference from continuous spatiotemporal evidence when asking individuals to make causal inferences about objects in 2D physical scenes where it is unlikely that participants will ever see exactly the same thing happen twice (Bass et al., 2021; Bramley, Gerstenberg, Tenenbaum et al., 2018; Ludwin-Peery et al., 2021; Ullman et al., 2018). Due to the high dimensionality of the clips used in these studies, it is crucial to leverage one's pre-existing mechanistic theory (e.g., a familiarity with everyday intuitive physical dynamics) to discover latent causal features such as objects' masses or force relations within the space of a short observation.

We argue that temporal evidence shares characteristics with both atemporal and spatiotemporal evidence. Like atemporal data, temporal evidence permits some discretization and aggregation, as effect events may occur multiple times without the necessity of having individual identifications (e.g., the bacteria culture example in Figure 1b; Gong & Bramley, 2024; Greville & Buehner, 2007; Griffiths & Tenenbaum, 2005; Pacer & Griffiths, 2012). This allows for type-level reasoning, about how the rate of effect occurrence changes after a putative cause occurs (i.e., the rate-based scheme). At the same time, temporal information also invites token-level reasoning. When one cause produces a very limited number of effect events (e.g., only one per component), the precise delays between each cause and effect and the prior expectations about causal and non-causal delays becomes important. Type-level causal conclusions will arise from the detailed inference about which specific occurrence of a cause was responsible for this specific occurrence of the effect (i.e., the event-based scheme). As such, the general Bayesian inference framework allows us to express whatever mechanistic or ontological commitments we believe capture a particular causal inference domain.

Computational cost concerns

Continuous time allows for precise temporal information, with each event having its unique time point and relationship with all other events. Events of different classes are often intermingled, and events of the same class may occur many times within the same observation. However, this precision and combinatorial credit assignment issue poses computational challenges and becomes infeasible when there are many events under consideration. Strictly, observing a causal system in continuous time with uncertainty about the true causal delays, any event could theoretically be the result of any event that happened in the past. As a real-world example there are diseases, such as the bovine variant of Creutzfeldt-Jakob disease, that have very long incubation periods. The cause of a disease onset could be traced back to something eaten 15 years ago (Valleron et al., 2001), but so many candidate events will occur within this period that it is impossible to consider them all. As such, a real cognizer should take seriously the trade-off between cost-of-computation and accuracy when reasoning about causal structure in their environment. The event-based and rate-based schemes we present here provide two levels at which one can process the same evidence, with the former generally more costly in its analysis of the micro-level delay details and the other a more abstracted and efficient way to capture the macro-level rate changes. By considering both approaches, a learner could flexibly choose or learn to represent a domain in a way that is sufficient and practical for their purposes. A rate-based scheme is especially useful when dealing with a large number of effects where it usually requires less computation. However, determining the appropriate granularity for rate calculation introduces another cost-benefit trade off that needs to be explored.

Abstraction and reduction: Moving between levels

We have focused on learning from events in continuous time, whereas other studies have examined causal learning from interactions with or observations of continuous valued variables varying in continuous time (Btesh et al., in press; Davis et al., 2020; Rehder et al., 2022; Soo & Rottman, 2018; Zhang & Rottman, 2023). Rather than viewing these as completely separate tasks, we think it is more fruitful to think of continuous and eventive representations as complementary ways of modeling and explaining causal phenomena. To illustrate this, consider the predator–prey relationship, such as that between lynx and hares. At a low level, we might model individual events such as a individual lynx catching and eating an individual hare. Abstracting this to a higher level, we might analyze how populations of lynx and hares change over time, based on their populations, or similarly on their fluctuating birth and predation rates. At a higher level, we can investigate how each species experiences cyclic patterns of population-scale events and shocks such as “bloom” and “collapse”, and analyze the progression of these an event representation again. By abstracting upwards or unpacking downwards, it seems that a flexible reasoner can cycle unboundedly between representations in terms of continuous values and those in terms of discretized events with the more appropriate choice determined by its utility in guiding action rather than adherence to metaphysical reality.

Importantly, it is not necessary to limit causal reasoning to the level natively provided by the data. When modeling data from Gong et al. (2023), we showed that the rate-based model outperformed the event-based model in capturing human performance in identifying cyclic structures, although the evidence was actually generated following an event-based scheme. This kind of abstraction can be boundedly rational. People may spontaneously abstract to a relatively continuous representation (i.e., the rate) when it makes computational sense to do so even if this prohibits some subtler mechanistic considerations. They might also do the reverse. Rehder et al. (2022) model structure inference in a setting involving continuous variables varying in real time. Their modeling suggests participants are overwhelmed by the full dynamics and rather abstract these into a handful macro-scale “events” – essentially treating moments of dramatic increase or decrease as events and performing token-level causal inference about relationships between these. Taken together, this all suggests that people have the ability to adopt computationally sensible representations. The layout of the event-based scheme in this paper also suggests that, contrary to the common assumption in computer science and other time-related cognitive models, the representation and operation of temporally relevant data are not necessarily bound to the discretization of time into bins, which may lead to inefficiency — representing numerous empty bins where no events occur — or distortion due to inappropriate bin width, where multiple events per bin reduce temporal accuracy and obscure order information. Reasoning from temporal evidence introduces a different and often more difficult computational challenge compared to the previously studied atemporal learning setting. As such, it can serve as a useful setting for studying the mechanisms that guide bounded rationality (Lieder & Griffiths, 2020; Simon, 1982).

Limitations and future directions

In the current paper, we treat events as instantaneous, occupying a single time point on the timeline. This means that a generative cause will always produce an observable effect even if that effect occurs close to another effect event. These point event representations could be seen as simplifications of everyday events that ignore their duration. Sometimes the duration of everyday events is too long to ignore: wet ground will stay wet for some time, tanned skin will fade slowly. In these cases, generative events can easily be overshadowed by already-occurring events (i.e., we might not notice that the sprinkler system came on because it had also been raining). This results in more complex causal inference scenarios such as preemption and over-determination (Gerstenberg et al., 2021; Lombrozo, 2010). These sorts of situations can be handled by extensions to the Poisson-Gamma framework, especially using event-based scheme, by incorporating the relevant mechanistic knowledge to the model (cf. Bramley, Gerstenberg, Mayrhofer et al., 2018). However, such situations are relatively unexplored in the temporal causal learning setting. When events have richer internal structure, such as a gradual onset and offset, the causal learning process becomes entangled with the question of people abstract continuous input into events. Future research could explore the possibility of integrating causal considerations with the theory of event segmentation (Altmann & Ekves, 2019) to build a comprehensive model of how discretized representations arise form continuous inputs.

Causal cognition researchers have used linear regression (Rottman, 2016; Soo & Rottman, 2018) and the Ornstein–Uhlenbeck (OU) process (Btesh et al., in press; Davis et al., 2020; Gong & Bramley, 2022; Rehder et al., 2022; Uhlenbeck & Ornstein, 1930) to generate continuous-variable, continuous-time dynamics and treated inference within the requisite model class as determining the normative solution to learn causal mechanisms. Although our event-level framework does not extend to continuous variables, linking it to this setting is a goal for future research. As an initial example, Gong and Bramley (2022) used the OU process to construct and test human inferences about continuous causal dynamics influences with a variety of properties. Similar to learning from event sequences, participants made stronger more confident judgments when the lag between changes to cause and effect variables' values was shorter and when the changes were more dramatic. This finding may suggest that combining the algorithms used in continuous-variable studies with the current Poisson-Gamma framework can help us better understand how people infer discrete event structure to explain continuous dynamics.

Although we focus on the role of time in causal reasoning, the causal relationships we investigate in this paper still align with the primary focus of atemporal causal learning literature: generation and prevention. In fact, causal relationships can be much richer once the time dimension is taken into account. For example, a cause can be "zero sum", in the sense of merely altering the timing of subsequent effects without affecting their frequency (Bennett, 1987). In other words, something might have a large causal effect on something else, not because it generates or prevents events but instead modifies *when* those events occur, by *hastening* or *delaying* them, or otherwise influencing their occurrence in time (Greville et al., 2020). Hastening and delaying are two relationships that are less studied in causal literature but extremely common in daily life; for example, we regularly experience things like transport and logistics delays. This richer space of causal difference-making could be studied from a signal detection perspective, by examining the conditions under which people notice when an event has had a causal influence on the occurrence of another, and when they can recognize what functional form that influence has taken. These kinds of situations could also be studied from a causal language perspective (Beller & Gerstenberg, in press; Wolff, 2007). For example, there might be important differences between someone judging that a poison killed A than that it hastened A's death. With the temporal dimension in play, further research can investigate detection and representation of a broader range of causal influence patterns both empirically and computationally.

In this paper we showed that participants' judgments across seen experiments can be unified under a rational framework. This is just one part of the project of understanding how people make these judgments. Humans have limited cognitive resources for receiving, processing, and memorizing information. These constraints significantly impact temporal causal learning, especially when the time gap between a cause and its effect is long. This paper focuses on developing the computational-level account, while it is important not to stop there but to use this to help investigate the process level (Marr, 1982). Some of the shifting computational demands of causal inference show up as the normative account is applied to different settings that scale differently in terms of the number of relevant events, variables and constraints on their potential relationships and functional forms. We might break process-level investigation into least two aspects. One aspect is understanding how temporal evidence is processed and compressed under our memory and computational constraints (Gallistel & Gibbon, 2000; Gong et al., 2023). The other aspect concerns how the causal hypothesis space and priors are constructed and searched (Bramley & Xu, 2023; Buchanan et al., 2010; Gong et al., 2024). This issue of understanding progressive search over causal hypotheses is especially important in the temporal setting given the strong now-or-never pressure on online computation relative to self-paced evidence setting. We hope that future studies can use this rational analysis of temporal causal structure induction to further explore human learning mechanisms, and how the computations involved interact with temporal scale (second, hours, days; Willett & Rottman, 2021).

A final point why understanding temporal causal reasoning is vital is that it not only drives the identification of causal mechanisms among known events but also determines *when* and *where* we direct our attention. For example, we can actively anticipate and look for events that are as yet unobserved but are predicted by the existence of our causal theories. This is a critical part of scientific practice: with a good mechanistic model researchers can decide when to measure the outcomes of their experiments, such as when a drug's influence on a person should be most apparent. Making these choices in a theory-guided way seems almost as important for causal discovery as the technique of random-assignment experiments, yet has received far less attention. Future studies should investigate how scientists and laypeople make observations or conduct experiments that consider the information measurable from the intermediate processes, as well as from the final outcome variables. By doing so, we can build a more comprehensive picture of scientific discovery as well as everyday cognition.

Conclusion

We inhabit a complex environment filled with continuous spatiotemporal causal dynamics. In order to form a practical causal understanding of this dynamic world, it seems essential for our minds to process temporal information effectively and efficiently. Despite fruitful empirical findings regarding how individuals behave in various time-based causal learning tasks, there is a lack of a unified theoretical framework to integrate behavioral predictions across all these tasks. In this paper, we present such a rational framework for causal induction based on the Poisson-Gamma statistical distribution families. We show how this framework aligns with human causal judgments. The framework grounds the basic philosophical intuitions about causality, and captures core qualitative empirical patterns that have long been seen in human learning studies. Quantitatively, the model is a good fit with human judgments across seven very different datasets. By laying out this framework, we take a key step towards understanding the computational task faced by humans and other agents when inducing a model of their environment. We hope the framework will serve as a benchmark for further investigation of the cognitive processes involved in generating and adapting causal representations, as well as how and why these may differ across different domains and timescales.

References

Ahn, W.-k., Kalish, C. W., Medin, D. L., & Gelman, S. A. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition*, *54*(3), 299–352.

Allan, L. G. (1980). A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society*, *15*(3), 147–149.

Almaatouq, A., Griffiths, T. L., Suchow, J. W., Whiting, M. E., Evans, J., & Watts, D. J. (2022). Beyond playing 20 questions with nature: Integrative experiment design in the social and behavioral sciences. *Behavioral and Brain Sciences*, 1–55.

Altmann, G., & Ekves, Z. (2019). Events as intersecting object histories: A new theory of event representation. *Psychological Review*, *126*(6), 817–840.

Anderson, J. R., & Sheu, C.-F. (1995). Causal inferences as perceptual judgments. *Memory & Cognition*, *23*(4), 510–524.

Anderson, J. R. (1990). *The adaptive character of thought*. Psychology Press.

Bass, I., Smith, K. A., Bonawitz, E., & Ullman, T. D. (2021). Partial mental simulation explains fallacies in physical reasoning. *Cognitive Neuropsychology*, *38*(7-8), 413–424.

Beller, A., & Gerstenberg, T. (in press). A counterfactual simulation model of causal language. *Psychological Review*.

Bennett, J. (1987). Event causation: The counterfactual analysis. *Philosophical Perspectives*, *1*, 367–386.

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.

Bramley, N. R., Gerstenberg, T., Mayrhofer, R., & Lagnado, D. A. (2018). Time in causal structure learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(12), 1880–1910.

Bramley, N. R., Gerstenberg, T., Mayrhofer, R., & Lagnado, D. A. (2019). Intervening in time. *Time and causality across the sciences*, 86–115.

Bramley, N. R., Gerstenberg, T., Tenenbaum, J. B., & Gureckis, T. M. (2018). Intuitive experimentation in the physical world. *Cognitive Psychology*, *195*, 9–38.

Bramley, N. R., Lagnado, D. A., & Speekenbrink, M. (2015). Conservative forgetful scholars: How people learn causal structure through sequences of interventions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(3), 708–731.

Bramley, N. R., Mayrhofer, R., Gerstenberg, T., & Lagnado, D. A. (2017). Causal learning from interventions and dynamics in continuous time. In G. Gunzelmann, A. Howes, T. Tenbrink & E. J. Davelaar (Eds.), *Proceedings of the 39th annual conference of the cognitive science society* (pp. 150–155).

Bramley, N. R., & Xu, F. (2023). Active inductive inference in children and adults: A constructivist perspective. *Cognition*, *238*, 105471.

Btesh, V. J., Bramley, N., Speekenbrink, M., & Lagnado, D. (in press). Less is more: Local focus in continuous time causal learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Buchanan, D., Tenenbaum, J., & Sobel, D. (2010). Edge replacement and nonindependence in causation. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd annual conference of the cognitive science society* (pp. 919–924).

Buehner, M. J., Cheng, P. W., & Clifford, D. (2003). From covariation to causation: A test of the assumption of causal power. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(6), 1119.

Buehner, M. J., & May, J. (2003). Rethinking temporal contiguity and the judgement of causality: Effects of prior knowledge, experience, and reinforcement procedure. *The Quarterly Journal of Experimental Psychology Section A*, *56*(5), 865–890.

Buehner, M. J., & May, J. (2004). Abolishing the effect of reinforcement delay on human causal learning. *Quarterly Journal of Experimental Psychology Section B*, *57*(2), 179–191.

Buehner, M. J., & McGregor, S. (2006). Temporal delays can facilitate causal attribution: Towards a general timeframe bias in causal induction. *Thinking & Reasoning*, *12*(4), 353–378.

Carroll, C., & Cheng, P. (2009). Preventative scope in causation. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31th annual conference of the cognitive science society* (pp. 833–838).

Cartwright, N. (1994). *Nature's capacities and their measurement*. Oxford University Press.

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

Chung, S.-H. (1965). Effects of delayed reinforcement in a concurrent situation 1. *Journal of the Experimental Analysis of Behavior*, *8*(6), 439–444.

Clarke, R. (1946). An application of the poisson distribution. *Journal of the Institute of Actuaries*, *72*(3), 481–481.

Coenen, A., Rehder, B., & Gureckis, T. M. (2015). Strategies to intervene on causal systems are adaptively selected. *Cognitive Psychology*, *79*, 102–133.

Courville, A. C., Daw, N. D., & Touretzky, D. S. (2006). Bayesian theories of conditioning in a changing world. *Trends in Cognitive Sciences*, *10*(7), 294–300.

Craik, K. J. W. (1967). *The nature of explanation* (Vol. 445). CUP Archive.

Davis, Z., Bramley, N. R., & Rehder, B. (2020). Causal structure learning in continuous systems. *Frontiers in Psychology*, *11*, 244.

Dean, T., & Kanazawa, K. (1989). A model for reasoning about persistence and causation. *Computational Intelligence*, *5*(2), 142–150.

Dennett, D. C. (1971). Intentional systems. *The Journal of Philosophy*, *68*(4), 87–106.

Eberhardt, F., Glymour, C., & Scheines, R. (2012). On the number of experiments sufficient and in the worst case necessary to identify all causal relations among n variables. *Proceedings of the Twenty-First Conference on Uncertainty in Artificial Intelligence*.

Eckstein, M. K., & Collins, A. G. (2020). Computational evidence for hierarchically structured reinforcement learning in humans. *Proceedings of the National Academy of Sciences*, *117*(47), 29381–29389.

Einhorn, H. J., & Hogarth, R. M. (1986). Judging probable cause. *Psychological Bulletin*, *99*(1), 3–19.

Fernando, C. (2013). From blickets to synapses: Inferring temporal causal networks by observation. *Cognitive Science*, *37*(8), 1426–1470.

Fernbach, P. M., & Sloman, S. A. (2009). Causal learning with local computations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(3), 678–693.

Fink, D. (1997). *A compendium of conjugate priors* (tech. rep.). https://www.johndcook.com/CompendiumOfConjugatePriors.pdf.

Gallistel, C. R. (2021). Robert rescorla: Time, information and contingency. *Revista de Historia de la Psicologıa*, *42*(1), 7–21.

Gallistel, C. R., Craig, A. R., & Shahan, T. A. (2014). Temporal contingency. *Behavioural Processes*, *101*, 89–96.

Gallistel, C. R., Craig, A. R., & Shahan, T. A. (2019). Contingency, contiguity, and causality in conditioning: Applying information theory and weber's law to the assignment of credit problem. *Psychological Review*, *126*(5), 761–773.

Gallistel, C. R., & Gibbon, J. (2000). Time, rate, and conditioning. *Psychological Review*, *107*(2), 289–344.

Gallistel, C. R., & Shahan, T. A. (2024). Time-scale invariant contingency yields one-shot reinforcement learning despite extremely long delays to reinforcement. *Proceedings of the National Academy of Sciences*, *121*(30), e2405451121.

Gallistel, C. R., & Wilkes, J. T. (2016). Minimum description length model selection in associative learning. *Current Opinion in Behavioral Sciences*, *11*, 8–13.

Garcia, J., Ervin, F. R., & Kölling, R. A. (1966). Learning with prolonged delay of reinforcement. *Psychonomic Science*, *5*(3), 121–122.

Gershman, S. J. (2015). A unifying probabilistic view of associative learning. *PLoS Computational Biology*, *11*(11), e1004567.

Gershman, S. J. (2017). Reinforcement learning and causal models. In M. Waldmann (Ed.), *The oxford handbook of causal reasoning* (pp. 295–306). Oxford University Press.

Gershman, S. J. (2024). Rate estimation revisited. *PsyArXiv*.

Gershman, S. J., Blei, D. M., & Niv, Y. (2010). Context, learning, and extinction. *Psychological Review*, *117*(1), 197.

Gershman, S. J., Norman, K. A., & Niv, Y. (2015). Discovering latent causes in reinforcement learning. *Current Opinion in Behavioral Sciences*, *5*, 43–50.

Gerstenberg, T., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2021). A counterfactual simulation model of causal judgments for physical events. *Psychological Review*, *128*(5), 936–975.

Gerstenberg, T., & Stephan, S. (2021). A counterfactual simulation model of causation by omission. *Cognition*, *216*, 104842.

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

Gong, T. (2023). *Causal induction in time* (Doctoral dissertation). University of Edinburgh.

Gong, T., & Bramley, N. R. (2020). What you didn't see: Prevention and generation in continuous time causal induction. In S. Denison, M. Mack, Y. Xu & B. Armstrong (Eds.), *Proceedings of the 42ed annual conference of the cognitive science society* (pp. 2908–2914).

Gong, T., & Bramley, N. R. (2022). Intuitions and perceptual constraints on causal learning from dynamics. In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Meeting of the Cognitive Science Society* (pp. 1455–1461).

Gong, T., & Bramley, N. R. (2023). Continuous time causal structure induction with prevention and generation. *Cognition*, *240*, 105530.

Gong, T., & Bramley, N. R. (2024). Evidence from the future. *Journal of Experimental Psychology: General*, *153*(3), 864–872.

Gong, T., Gerstenberg, T., Mayrhofer, R., & Bramley, N. R. (2023). Active causal structure learning in continuous time. *Cognitive Psychology*, *140*, 101542.

Gong, T., Valentin, S., Lucas, C. G., & Bramley, N. R. (2024). Paradoxical parsimony: How latent complexity favors theory simplicity. In L. Samuelson, S. Frank, M. Toneva, A. Mackey & E. Hazeltine (Eds.), *Proceedings of the 46th annual conference of the cognitive science society* (pp. 2428–3434).

Goodman, N. (1983). *Fact, fiction, and forecast*. Harvard University Press.

Grabenhorst, M., Maloney, L. T., Poeppel, D., & Michalareas, G. (2021). Two sources of uncertainty independently modulate temporal expectancy. *Proceedings of the National Academy of Sciences*, *118*(16), e2019342118.

Grabenhorst, M., Michalareas, G., Maloney, L. T., & Poeppel, D. (2019). The anticipation of events in time. *Nature Communications*, *10*(1), 5802.

Greville, W. J., & Buehner, M. J. (2007). The influence of temporal distributions on causal induction from tabular data. *Memory & Cognition*, *35*(3), 444–453.

Greville, W. J., & Buehner, M. J. (2010). Temporal predictability facilitates causal learning. *Journal of Experimental Psychology: General*, *139*(4), 756–771.

Greville, W. J., & Buehner, M. J. (2016). Temporal predictability enhances judgements of causality in elemental causal induction from both observation and intervention. *Quarterly Journal of Experimental Psychology*, *69*(4), 678–697.

Greville, W. J., Buehner, M. J., & Johansen, M. K. (2020). Causing time: Evaluating causal changes to the when rather than the whether of an outcome. *Memory & Cognition*, *48*, 200–211.

Griffiths, T. L. (2020). Understanding human intelligence through human limitations. *Trends in Cognitive Sciences*, *24*(11), 873–883.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology*, *51*(4), 334–384.

Griffiths, T. L., & Tenenbaum, J. B. (2007). From mere coincidences to meaningful discoveries. *Cognition*, *103*(2), 180–226.

Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological Review*, *116*(4), 661–716.

Hagmayer, Y., & Waldmann, M. R. (2002). How temporal assumptions influence causal judgments. *Memory & Cognition*, *30*(7), 1128–1137.

Halpern, J. Y. (2016). *Actual causation*. MIT Press.

Hamou, N., Gershman, S. J., & Reddy, G. (2025). Reconciling time and prediction error theories of associative learning. *bioRxiv*, 2025–01.

Hume, D. (1740). *A treatise of human nature*. Oxford University Press (2000 reprint).

Icard, T. F., Kominsky, J. F., & Knobe, J. (2017). Normality and actual causal strength. *Cognition*, *161*, 80–93.

Kalmbach, A., Chun, E., Taylor, K., Gallistel, C. R., & Balsam, P. D. (2019). Time-scale-invariant information-theoretic contingencies in discrimination learning. *Journal of Experimental Psychology: Animal Learning and Cognition*, *45*(3), 280–289.

Kim, N. S., & Ahn, W.-k. (2002). Clinical psychologists' theory-based representations of mental disorders predict their diagnostic reasoning and memory. *Journal of Experimental Psychology: General*, *131*(4), 451–476.

Kolvoort, I., Davis, Z. J., Rehder, B., & van Maanen, L. (2025). Models of variability in probabilistic causal judgments. *Computational Brain & Behavior*, *8*, 162–188.

Lagnado, D. A., & Sloman, S. (2002). Learning causal structure. In W. Gray & C. Schunn (Eds.), *Proceedings of the 24th annual meeting of the cognitive science society* (pp. 560–565).

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

Lagnado, D. A., & Sloman, S. A. (2006). Time as a guide to cause. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(3), 451–460.

Lagnado, D. A., & Speekenbrink, M. (2010). The influence of delays in real-time causal learning. *The Open Psychology Journal*, *3*(1), 184–195.

Lagnado, D. A., Waldmann, M. R., Hagmayer, Y., & Sloman, S. A. (2007). Beyond covariation. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 154–172). Oxford University Press.

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, *40*, e253.

Lauritzen, S. L., & Richardson, T. S. (2002). Chain graph models and their causal interpretations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *64*(3), 321–348.

Libet, B. (2009). *Mind time: The temporal factor in consciousness*. Harvard University Press.

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, *43*, 1–60.

Lombrozo, T. (2010). Causal–explanatory pluralism: How intentions, functions, and mechanisms influence causal ascriptions. *Cognitive Psychology*, *61*(4), 303–332.

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.

Luce, R. D. (1959). *Individual choice behavior*. Wiley.

Ludwin-Peery, E., Bramley, N. R., Davis, E., & Gureckis, T. M. (2021). Limits on simulation approaches in intuitive physics. *Cognitive Psychology*, *127*, 101396.

Malthus, T. R. (1872). *An essay on the principle of population*. Reeves & Turner.

Markant, D. B., & Gureckis, T. M. (2014). Is it better to select or to receive? learning via active and passive hypothesis testing. *Journal of Experimental Psychology: General*, *143*(1), 94–122.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. MIT Press.

Mendelson, R., & Shultz, T. R. (1976). Covariation and temporal contiguity as principles of causal inference in young children. *Journal of Experimental Child Psychology*, *22*(3), 408–412.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., et al. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529–533.

Myerson, J., & Green, L. (1995). Discounting of delayed rewards: Models of individual choice. *Journal of the Experimental Analysis of Behavior*, *64*(3), 263–276.

Nikolic, M., & Lagnado, D. A. (2015). There aren't plenty more fish in the sea: A causal network approach. *British Journal of Psychology*, *106*(4), 564–582.

O'Neill, K., Henne, P., Bello, P., Pearson, J., & De Brigard, F. (2022). Confidence and gradation in causal judgment. *Cognition*, *223*, 105036.

O'Neill, K., Henne, P., Pearson, J., & De Brigard, F. (2024). Modeling confidence in causal judgments. *Journal of Experimental Psychology: General*, *153*(8), 2142–2159.

Pacer, M. (2016). *Mind as theory engine: Causation, explanation and time* (Doctoral dissertation). UC Berkeley.

Pacer, M., & Griffiths, T. L. (2012). Elements of a rational framework for continuous-time causal induction. In N. Miyake, D. Peebles & R. P. Cooper (Eds.), *Proceedings of the 34th annual conference of the cognitive science society* (pp. 833–838).

Pacer, M., & Griffiths, T. L. (2015). Upsetting the contingency table: Causal induction over sequences of point events. In D. C. Noelle, R. Dale, A. S. Warlaumont,

J. Yoshimi, T. Matlock, C. D. Jennings & P. P. Maglio (Eds.), *Proceedings of the 37th annual conference of the cognitive science society* (pp. 1805–1810).

Pearl, J. (2000). *Causality*. Cambridge University Press (2009 reprint).

Perales, J. C., & Shanks, D. R. (2007). Models of covariation-based causal judgment: A review and synthesis. *Psychonomic Bulletin & Review*, *14*, 577–596.

Pishro-Nik, H. (2014). *Introduction to probability, statistics, and random processes*. Kappa Research, LLC Blue Bell, PA, USA.

Rehder, B. (2017). Reasoning with causal cycles. *Cognitive Science*, *41*, 944–1002.

Rehder, B., Davis, Z. J., & Bramley, N. (2022). The paradox of time in dynamic causal systems. *Entropy*, *24*(7), 863.

Rescorla, R. A. (1968). Probability of shock in the presence and absence of cs in fear conditioning. *Journal of Comparative and Physiological Psychology*, *66*(1), 1–5.

Rescorla, R. A., & Wagner, A. R. (1972). A theory on pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. Black & W. Prokasy (Eds.), *Classical conditioning ii: Current theory and research* (pp. 64–99). Appleton Century Crofts.

Ross, L. (2024). Causal constraints in the life and social sciences. *Philosophy of Science*, *91*(5), 1068–1077.

Rottman, B. M. (2016). Searching for the best cause: Roles of mechanism beliefs, autocorrelation, and exploitation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(8), 1233.

Rottman, B. M., & Hastie, R. (2014). Reasoning about causal relationships: Inferences on causal networks. *Psychological Bulletin*, *140*(1), 109–139.

Rottman, B. M., & Keil, F. C. (2012). Causal structure learning over time: Observations and interventions. *Cognitive Psychology*, *64*(1-2), 93–125.

Scheines, R. (1997). *An introduction to causal inference*. Carnegie Mellon University.

Schlottmann, A., Cole, K., Watts, R., & White, M. (2013). Domain-specific perceptual causality in children depends on the spatio-temporal configuration, not motion onset. *Frontiers in Psychology*, *4*, 365.

Schultz, W. (2015). Neuronal reward and decision signals: From theories to data. *Physiological Reviews*, *95*(3), 853–951.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593–1599.

Shanks, D. R., & Dickinson, A. (1991). Instrumental judgment and performance under variations in action-outcome contingency and contiguity. *Memory & Cognition*, *19*(4), 353–360.

Shanks, D. R., Pearson, S. M., & Dickinson, A. (1989). Temporal contiguity and the judgement of causality by human subjects. *The Quarterly Journal of Experimental Psychology*, *41*(2), 139–159.

Simon, H. A. (1982). *Models of bounded rationality: Empirically grounded economic reason.* MIT press.

Sloman, S. A. (2005). *Causal models: How people think about the world and its alternatives.* Oxford University Press.

Sloman, S. A., & Lagnado, D. A. (2005). Do we "do"? *Cognitive Science*, *29*, 5–39.

Sloman, S. A., Love, B. C., & Ahn, W.-K. (1998). Feature centrality and conceptual coherence. *Cognitive Science*, *22*(2), 189–228.

Soo, K. W., & Rottman, B. M. (2018). Causal strength induction from time series data. *Journal of Experimental Psychology: General*, *147*(4), 485–513.

Stephan, S., Mayrhofer, R., & Waldmann, M. R. (2020). Time and singular causation–a computational model. *Cognitive Science*, *44*(7), e12871.

Stephan, S., Placì, S., & Waldmann, M. R. (2021). Evaluating general versus singular causal prevention. In T. Fitch, C. Lamm, H. Leder & K. Teßmar-Raible (Eds.), *Proceedings of the 43th annual conference of the cognitive science society* (pp. 1402–1408).

Steyvers, M., Tenenbaum, J. B., Wagenmakers, E.-J., & Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive Science*, *27*(3), 453–489.

Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science*, *12*(1), 49–100.

Tarpy, R. M., & Sawabini, F. L. (1974). Reinforcement delay: A selective review of the last decade. *Psychological Bulletin*, *81*(12), 984–997.

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences*, *10*(7), 309–318.

Uhlenbeck, G. E., & Ornstein, L. S. (1930). On the theory of the brownian motion. *Physical Review*, *36*(5), 823–841.

Ullman, T. D., Stuhlmüller, A., Goodman, N. D., & Tenenbaum, J. B. (2018). Learning physical parameters from dynamic scenes. *Cognitive Psychology*, *104*, 57–82.

Valentin, S., Bramley, N. R., & Lucas, C. G. (2020). Learning hidden causal structure from temporal data. In S. Denison, M. Mack, Y. Xu & B. Armstrong (Eds.), *Proceedings of the 42ed annual conference of the cognitive science society* (pp. 1906–1912).

Valentin, S., Bramley, N. R., & Lucas, C. G. (2022). Discovering common hidden causes in sequences of events. *Computational Brain & Behavior*, 1–23.

Valleron, A.-J., Boelle, P.-Y., Will, R., & Cesbron, J.-Y. (2001). Estimation of epidemic size and incubation time based on age characteristics of vcjd in the united kingdom. *Science*, *294*(5547), 1726–1728.

Waldmann, M. R., & Hagmayer, Y. (2005). Seeing versus doing: Two modes of accessing causal knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(2), 216–227.

Wang, J. X. (2021). Meta-learning in natural and artificial intelligence. *Current Opinion in Behavioral Sciences*, *38*, 90–95.

Willett, C. L., & Rottman, B. M. (2021). The accuracy of causal learning over long timeframes: An ecological momentary experiment approach. *Cognitive Science*, *45*(7), e12985.

Wolff, P. (2007). Representing causation. *Journal of Experimental Psychology: General*, *136*(1), 82–111.

Woodward, J. (2021). *Causation with a human face: Normative theory and descriptive psychology*. Oxford University Press.

Wu, M., & Cheng, P. W. (1999). Why causation need not follow from statistical association: Boundary conditions for the evaluation of generative and preventive causal powers. *Psychological Science*, *10*(2), 92–97.

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

Zhang, Y., & Rottman, B. M. (2023). Causal learning with interrupted time series data. *Judgment and Decision Making*, *18*, e30.

Zhang, Y., & Rottman, B. M. (2024). Causal learning with delays up to 21 hours. *Psychonomic Bulletin & Review*, *31*(1), 312–324.

Appendix A

Prevention Causation

The event-based scheme

Considering preventative causation generally increases the space of possible explanations and would hence affect the likelihood calculation for specific actual pathways. Concretely, for each observed effect e, we must jointly evaluate the probability that (1) it was caused by its presumed generative cause event g as well as that (2) it was not prevented by any of the set of presumed preventative cause events $\mathbf{p}$. Hidden (expected but unobserved) effects also contribute to the likelihood wherever we do not observe an expected effect of a generative cause. This could be due to (1) the generative cause failing to produce that effect, (2) that effect being prevented, or (3) that effect not having occurred yet:

$$\begin{aligned} P(\mathbf{z}|s;\mathbf{w}) = \prod_{g\to e\in\mathbf{z}} & \underbrace{w_g \cdot P_d(t_e - t_g|\alpha,\beta)\cdot(1-P_p(e))}_{\text{Observed effects must have been generated and not prevented}} \\ \times \prod_{g\to h\in\mathbf{z}} & \underbrace{(1-w_g) + w_g\cdot P_d(t_h > t_{end}|\alpha,\beta) + w_g\cdot P_p(h)}_{\text{Unobserved expected effects must have failed or been prevented, or be still-to-occur}} \end{aligned} \quad \text{(A1)}$$

The event-based scheme provides also allows for flexibility in dealing with preventative causation $P_p(e)$ (the probability that e should have been prevented) based on different mechanisms of prevention. For instance, a preventative cause might block an effect from occurring at all for a specific time window. Alternatively, it might block the subsequent N effects from occuring before being "used up". A preventative cause might block all effects indiscriminately (e.g., operate on the effect variable), or selectively block effects from a particular cause (e.g., operate on the edge between two variables; Carroll & Cheng, 2009; Gerstenberg & Stephan, 2021).

The rate-based scheme

When a cause is generative, we expect the rate of its effect to temporarily increase, whereas we expect preventative causes to temporarily decrease the rate of their effects.

Intuitively, a preventative causal influence can be thought of as defeating some of the effects that would otherwise have occurred, meaning that it will have a proportional effect on the rate. As such, we assume preventative causes decrease the effect rate by a proportion ranging from 0 to a maximum level of ξ $(0 < \xi < 1)$. A preventative influence can also follow an incubation-decay process and be represented by a function of time $f(\xi, t) = \xi \cdot \frac{P_d(t|\alpha,\beta)}{P_d(\frac{\alpha-1}{\beta}|\alpha,\beta)}$.

This means preventative causation can be viewed as "thinning" processes that selectively filter out some effect events with a probability of ξ'. This contrasts with the natural way to think of generative causation as "superposition" where more events are added to the timeline. Combining multiple causes with a base rate of λ_0, the expected effect rate $f(\lambda, t)$ at the time unit t can be represented similar to the noisy-OR and noisy-AND-NOT principles by accounting for superposition and thinning as follows:

$$f(\lambda, t) = (\lambda_0 + \sum_{i \in \mathbf{g}} f(\lambda_i, t)) \prod_{j \in \mathbf{p}} (1 - f(\xi_j, t)) \tag{A2}$$

Appendix B

Time-scale Invariance Simulation

To demonstrate when the time-scale invariance property applies, we synthesize three conditions, each with causal delays $m_u = \{10, 12, 14, 16, 18\}$. The "no-match" condition follows the procedure in main text to use a fixed observation duration of 300 time units, with $w_c = 1$, $k_b = 3$, and $i_u = 1$ (similar to the main text we apply the number of cause events $k_c = 5$ for the event-based model and $k_c = 10$ for the rate-based model). In the "baserate-match" condition, the observation duration is scaled according to the causal delay. We retain 300 time units when $m_u = 10$, but adjust to 360, 420, 480, and 540 time units for the other delays, respectively. Causal events and base rate events are sampled within these new observation durations, ensuring that each causal delay has the same number of observed events (k_c and k_b), while the rate of background effect events scales with the causal delay. Specifically, for $m_u = 10$ the baserate is 3 per 300 time units (an average delay between baserate effects of 75 time units), and for $m_u = 12$ the baserate adjusts to 3 per 360 time units (with an average delay between base rate effects of 90 time units). The prior is also scaled accordingly. For example, if the observation duration is 360, we use $\mu \sim U(0, 360)$ and $\mu_b \sim U(0, 360)$ instead. In the "all-match" condition, we further scale the delay variance by setting $i_u = \{1, 1.2, 1.4, 1.6, 1.8\}$ for each causal delay respectively.

As shown in Figure B1, the tendency to favor a causal structure can remain at the same level even when the causal delay increases, as long as the environment is time-scale invariant (i.e., the observation duration, the base rate, and the delay variance are all scaled/matched according to the length of the causal delays). The tendency to favor the causal hypothesis decreases as the causal delay increases if the contextual factors do not scale, and it increases with the causal delay if the baserate-relevant factors are scaled but the delay variance is not.

The other intuitive way to think why the time-scale invariance would exist is to think, if we change the length scale of an event sequences from minutes to hours (and then the baserate will change from k per minute to k per hour). Accordingly, to create equally weak priors, we can for example, apply a uniform prior from 0 to 100 minutes (assuming the upper bound here is larger than the practically possible causal delay) in the short-delay condition and from 0 to 100 hours in the long-delay condition. In this case, the time unit becomes the only difference all calculations, but this does not matter because the time unit is pre-set arbitrarily before any calculation.

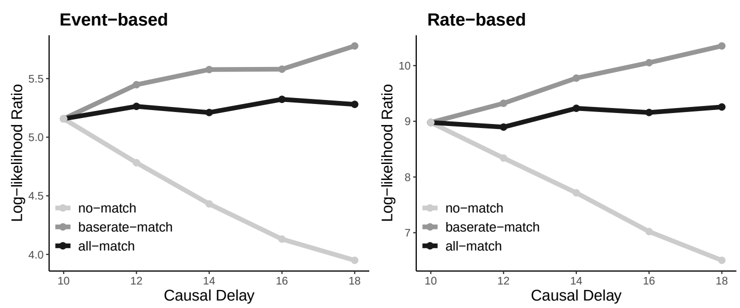


Figure B1

The log-likelihood ratio changes given the causal delays. A ratio above zero indicates that the model favors S_1 (the causal structure) over S_0 (the base rate structure).

Appendix C

The Persistent Effects of Delay Variance

Greville and Buehner (2010) found that the preference for unvaried delays persisted when the learning duration was doubled from 2 min to 4 min, which cannot be well explained by a difference in learning rates under the associative learning model (Chung, 1965; Rescorla, 1968). We here show that this result is easily captured by the Bayesian framework. We simulate different learning durations [100, 150, 200] and set the cause events to occur every 25 time units on average, with baserate effect events occurring every 50 time units on average. Each cause event generates an effect event with a delay sampled from $U(m_u - i_u, m_u + i_u)$ and a probability of w_c. As shown in Figure C1, when the average delay length m_u and the causal probability w_c are fixed, the effect of delay variance does not decrease as the learning duration increases. It instead increases as the evidence accumulates.

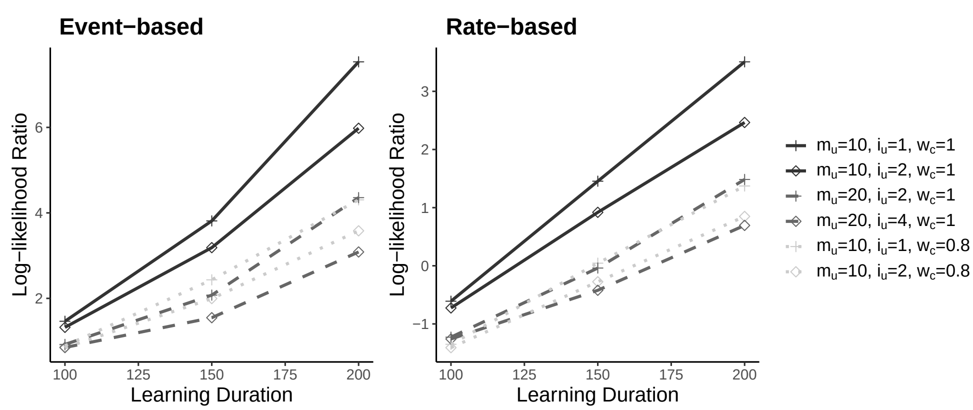


Figure C1

The log-likelihood ratio changes given the learning duration. A ratio above zero indicates that the model favors S_1 (the causal structure) over S_0 (the base rate structure).