Less is more: Local focus in continuous time causal learning

Abstract

In this study, we investigated human causal learning in a continuous time and space setting (a demo can be found here). We find participants to be capable active causal structure learners, and with the help of computational modelling explore how they mitigate the complexity of continuous dynamics data to achieve this. We propose that participants combine systematic interventions with a narrowed focus on causal dynamics that occur during and directly downstream of their interventions. This task-decomposition approach achieves comparable accuracy to attending to all the dynamics, while discarding almost half of the data. We argue this strategy makes sense from a resource rationality perspective: ignoring dynamics outside of interventions saves computational cost while the interventions naturally decompose the global learning problem into a series of more manageable sub-problems. We also find that when the causal relata are given real world labels, participants will use their domain specific priors to guide their structure inferences. In particular, individuals with accurate prior expectations were less likely to make the common local computations error of mistaking an indirect for a direct relationship. Overall, our experiments reinforce the idea that humans are frugal and intuitive active learners who combine actions and inference to optimize learning while minimizing effort.

Keywords active learning, causal reasoning, Bayesian inference, continuous time, domain-specific priors

1 Introduction

Consider an anaesthetist monitoring a patient in heart surgery. They can hear the patient's breathing and see their blood pressure in real time. Suddenly, the beeps of the heart monitor start to slow. The anaesthetist decides to intervene. They up the dose of atropine; the heart rate starts to rise, but other vitals plummet.

What caused this drop? Whatever the anaesthetist infers, it will depend on the dynamics observed recently about the patient, but also on previous knowledge about how the medication and the vitals causally relate. Medical care is one of many activities, e.g. stock trading or weather forecasting, which requires causal reasoning about abstract variables in continuous time. More generally, beliefs about cause and effect are a fundamental part of human cognition. They form the building blocks of our structured and predictive inner models of the world. Almost from

birth, we actively learn about our environment by experimenting upon it and reason about resulting evidence in ways that reflect inference to hidden causal laws and forces (Gopnik et al., 2004; Valentin et al., 2023).

Advances in statistical causal inference in the past few decades have yielded substantial insights into human causal learning and reasoning (see ~~Holyoak and Cheng, 2011; Bramley et al., 2019; Sloman and Lagnado, 2015~~ Waldmann, 2017; Bramley et a for reviews). Notably, causal Bayesian networks, introduced by Pearl (2009), provide a powerful framework for representing probabilistic causal beliefs that explain patterns of statistical dependence in terms of an underlying causal mechanism. The majority of studies in causal cognition make use of Pearl's formalism as the normative benchmark (some examples include Griffiths and Tenenbaum, 2009; Waldmann and Hagmayer, 2001; Tenenbaum et al., 2006; Lagnado et al., 2007; Hagmayer et al., 2010). However, it is premature to view behaviour through this single theoretical lens. In particular, standard causal Bayesian networks are designed to capture causal inference from and about static contingencies, missing out on dynamics which operate continuously in time. Causal Bayesian networks can also only represent directed acyclic causal systems and are thus most suitable for modelling systems with no feedback loops or cycles. This has led the study of more complex and dynamic causal learning to be slower to develop. Participants' tendency to violate assumptions of causal Bayesian networks such as the Markov condition (Rehder, 2014), explaining away (Rehder and Waldmann, 2017), zero-sum reasoning (Pilditch et al., 2019) or base-rate neglect (Stengård et al., 2022) have led to the emergence of alternative frameworks. While some used dynamic Bayesian networks and chain graphs to look at how participants learn cyclic structure (Rehder, 2017), their use remains limited. Bramley et al. (2017a), Davis et al. (2018) and Gong et al. (2023), have recently addressed people's ability to learn complex causal structures. Allowing causal loops and working in continuous time, this research has shown that, in spite of recurring biases, humans are capable learners in continuous time contexts.

In this paper, we are concerned with *active* causal learning, that is, learning which results from acting in the world with the purpose of gaining information (Coenen et al., 2019). The observations we learn from are often contingent on our own actions and thus also the beliefs which guided our choice of policy. This can be a blessing, allowing us to aim our actions where our uncertainty is higher, hence improving learning (Settles, 2012). It can also be a curse, discouraging us from sampling information about what we already know, even when it is wrong (Bramley et al. (2017a)). Active learning is thus about choosing an action and making inferences from its consequences. The former requires specifying an appropriate value function for future actions and is usually modelled as the expected information gained: the expected reduction in entropy in the posterior distribution over the parameters of interest (Settles, 2012; Gong et al., 2023). A causal active learner chooses interventions: actions which remove incoming causal links to a variable of the system, thus allowing for controlled investigation of its downstream effects (Pearl, 1997). In this work, we focus mostly on the latter part of active learning: making inferences from the observations that our interventions produced. Up until now, research on active causal learning has mostly ignored the role of prior knowledge. By focusing on learning about abstract structures, often with generic (e.g. A, B, C) variable names, application of prior knowledge was discouraged. However, in reality, we learn about meaningful causal structures not solely from observation, but also by using prior knowledge ~~about the variables~~

~~(Griffiths and Tenenbaum, 2009). People can~~ to make inferences about causal relations ~~prior to observing data~~ before observing data (Griffiths and Tenenbaum, 2009).For instance, if you know that the police's role is to enforce the law and that criminality is defined as breaking it, you may make the inference that increasing police numbers will reduce crime rates, before observing any policeman arrest a criminal. Consequently, informative prior beliefs are likely to interact in important ways with how people learn structure from evidence.

In the present study, we investigate the inference mechanisms that people use to learn causal structures relating continuous variables in continuous time, and how this learning is influenced by prior beliefs.

1.1 Causal structure learning and local computations

The computational complexity of causal structure learning in real-world settings, where the environment includes a seemingly infinite number of variables with unknown interactions, renders direct Bayesian inference implausible. Natural environments admit descriptions involving potentially infinite number of variables with complex relationships. As such, it seems plausible that causal structure learning relies greatly on heuristics and approximations.

As one such potential heuristic, Fernbach and Sloman (2009) propose that people infer causal structure through a Local Computation (LC) process, whereby they focus on the pairwise association between variables, ignoring interactions with the surrounding structure. This kind of approach generalises the single effect learning theory (Waldmann et al., 2008) which also proposed rats and humans learn causal structures by isolating each causal link. This broadly falls under the umbrella of task decomposition strategies, i.e. breaking down a learning problem into several sub-problems to simplify it and make gains in computational cost, usually at the expense of accuracy (Correa et al., 2022; Lieder and Griffiths, 2020). For instance, when people learn about causal relations between three variables (A, B and C), they might deconstruct this problem into six different sub-problems, of identifying the existence or non-existence of each potential directed connection between a pair of variables. Solutions to the sub-problems can then be combined in a final inference about the overall structure. LC assumes that each pair of variables (e.g. A and B) is analysed separately and treated as independent from the other variables when considering the causal link between them ($A \rightarrow B$). This is more computationally efficient compared to considering the much larger set of three variable structures directly, but results in systematic errors wherever multiple pathways can explain the evidence. A important prediction of the LC model is that people make errors wherever there are indirect effects. This prediction is best exemplified by a well-documented chain bias (Ahn and Dennis, 2000; Fernbach and Sloman, 2009; Bramley et al., 2015; Davis et al., 2020) whereby participants, when tasked to identify the structure of what is actually a $A \rightarrow B \rightarrow C$ chain, frequently report the true connections but also an incorrect additional direct link from A to C (Figure 1 A.). While LC can result in additional connections, this need not be the case. Figure 1 C. shows a chain with a direct negative link from A to C, which dampens the indirect effect going through B. In this instance, a LC learner will fail to report the direct link.

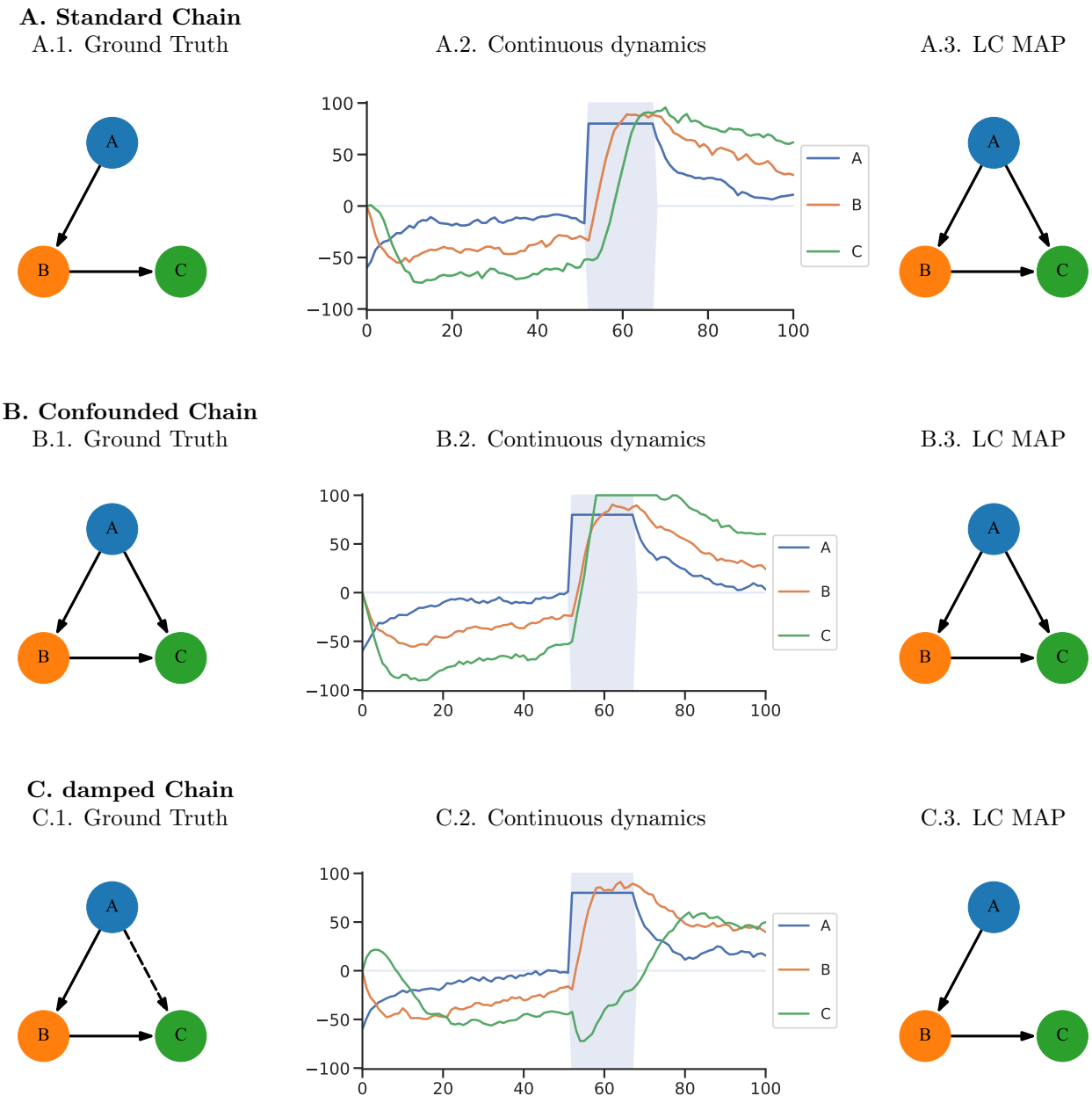


Figure 1: Ground truths, continuous time dynamics generated using Orstein-Uhlenbeck networks (Davis et al., 2020) with an intervention on the root node (blue shaded area) and the maximum a posteriori given the plotted data of a Local Computations model for A. a chain model, B. its confound and C. an inverted confound structure where the direct link is damped by an indirect effect. Full lines indicate positive links and dashed line negative links.

As a related approach, Bramley et al. (2017a) propose that to limit the computational resources needed to integrate new evidence, agents condition their next intervention and resulting observations on their currently held beliefs about the ground truth structure. This leads to a special case of local computations where instead of treating evidence for the presence of a link between two variables as (wrongly) coming from the marginal distribution, as LC

would predict, they treat it as coming from the conditional distribution of the variables given their current beliefs about the rest of the structure, allowing them to avoid the LC error if they already inferred the $A \rightarrow B$ and $B \rightarrow C$ connections) but not otherwise.

Recent work in continuous time by Rehder et al. (2022) showed that participants interacting with sluggish systems (where the continuous time dynamics played out more slowly) were less likely to make errors with indirect effects but this came at the cost of a higher overall error rate. While the sluggish dynamics made it easier to distinguish direct and indirect influences they also reduced the diagnosticity of the evidence overall. To explain these response patterns, Rehder et al. (2022) propose a model that identified links by discretising the continuous evidence into countable causal events, a process that was facilitated when dynamics are slowed.

1.2 Active causal structure learning

Humans not only learn from passive observation, but are also able to actively intervene on systems of interest in their environments to resolve uncertainty about their causal mechanisms (Rothe et al., 2018; Lagnado and Sloman, 2002, 2004, 2006; Steyvers et al., 2003; Rottman and Keil, 2012; Bramley et al., 2018; Davis et al., 2020). In Pearl's do-calculus, interventions, denoted $do(X = x)$, can be modelled as the learner fixing variable(s) in the system, for example setting X at a value x, such that $P(X = x) = 1$ without affecting X's parents. This is achieved graphically by temporarily changing the underlying model (see Figure 2) during the intervention to disconnect the intervened on variable from its normal causes (Gopnik et al., 2004; Steyvers et al., 2003). When provided with the ability to intervene in contingency settings, participants have been shown to make generally effective use of them. Lagnado and Sloman (2004) proposed that people think of interventions as strong temporal order cues for causal inference, guaranteeing that the intervened on variable occurred *before* anything else that changed. Bramley et al. (2017b) and Bramley et al. (2017a) showed that successful participants tend to spread out their interventions in time allowing for successive periods of observation of the consequences of their actions.

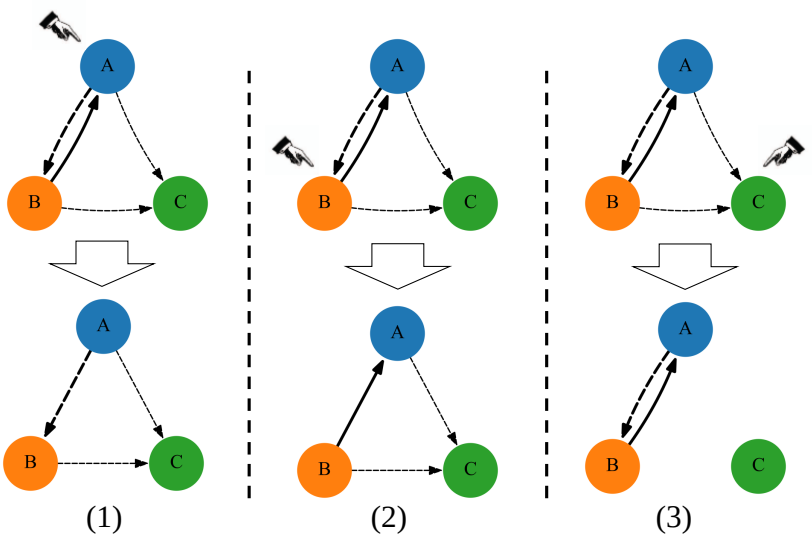


Figure 2: Intervening on causal systems using the *do* operator, as represented by the hand icon, changes their structure by cutting all incoming links to the variable intervened upon. Case 1 shows how intervening on A breaks a feedback loop between A and B. Case 2 shows the reverse. Case 3 shows that C is only an effect of both A and B as it becomes unaffected by them given an intervention. Full lines indicate positive links and dashed line negative links. Wider arrows indicate strong links and thinner arrow weak links.

In discrete time causal learning experiments, participants are not faced with an abundance of noisy data: on the contrary, they pick an intervention, and the subsequent observation reveals its consequences. ~~These settings mask any selection process about which data to attend to that may be needed in our daily lives where~~ However, the world continuously changes around us whether we act on it or not ~~. This discrepancy has been characterized as an ecological validity limitation for many experiments examining causal cognition~~ and, by discretising time so drastically, these settings obscure the cognitive processes which we must rely on to grapple with these dynamics (Bramley et al., 2019). For instance, selecting which data to attend to becomes a more urgent issue, as any informative set may not stay available for long, whether we intervene or not. Moving to a more naturalistic setting where variables are continuously valued and change continuously in time however presents a double challenge for learners and modellers, as the space of potential interventions goes from a handful of discrete actions to a continuous space of actions that can be extend in, and evolve over, continuous time. We believe that studying human inference in a continuously varying setting warrants tackling this complexity for this very reason, as behavioural control in daily life must engage with this apparently intractable problem.

In related work, Gong et al. (2023) focused on causal learning in continuous time, but restricted to discrete events evidence rather than fully continuous time series. They show that, under a resource-rational analysis (Lieder and Griffiths, 2020), participants interventions balanced consideration of expected information against anticipation of the computational costs of processing the evidence. Participants intervened on the causal systems they were learning about in ways that, while informative, would also control and often reduce the number of events occurring to generate more interpretable causal dynamics. In continuous time and continuously valued space however, such a formal model

has yet to emerge. Bramley et al. (2018) provide a descriptive account of interventions in continuous time and space settings, showing that participants' actions in an environment controlled by a realistic physics simulator, had some features of controlled experimentation. Participants appeared to take actions that exaggerate whatever environmental dynamics were most useful for revealing the property they were tasked with learning about while minimizing the confounding influence of other features of the environment. Rehder et al. (2022) showed that participants adaptively change their intervention strategies to suit the constraints of the experimental context, e.g. intervened for longer when there was a larger lag between cause and effect. They also showed that in cyclic models, through targeted interventions, participants were able to temporarily break loops (see Figure 2 for an example) and simplify causal structures. Their causal event abstraction model proposed that participant hold a value threshold which, if crossed by a variable during an intervention (see Figure 1 A.2., B.2. and C.2.), increments an evidence accumulation process for the existence of a link. A link from A to B is then inferred if B is frequently above the threshold during an intervention that holds A far from its equilibrium position. Crucially, they showed that this simple approach, which solely focuses on data generated during interventions, outperformed a standard LC account in capturing participants' judgements. This, in line with core principles of instrumental conditioning (Grice, 1948) and the proposal by Lagnado and Sloman (2004) that people use interventions as cues for inference, implies that combining action with focused attention may be a key feature of human learning in continuous time.

Drawing all these threads together we believe that in a continuous time learning setting where evidence is abundant but complex, a narrowed focus may be an adaptive strategy for learning, allowing learners to prioritize and target their computational resources to interpreting the most diagnostic dynamics. In short, we propose the strategy embodied by the local computations model (Fernbach and Sloman, 2009), can be generalized to the continuous setting along two dimensions, both anchored to the learner's interventions. First, there is *temporal locality* whereby we hypothesize that people focus their attention on the dynamics following their interventions (Lagnado and Sloman, 2004; Rehder et al., 2022). Second, there is *structural locality* where we propose that people focus on identifying the direct effects of the intervened-on variable rather than the full causal structure (Lagnado and Sloman, 2002; Waldmann et al., 2008).

1.3 Priors and domain-specificity

The role of priors in learning and judgments is not a new question in causal cognition. Tenenbaum et al. (2006) and Griffiths and Tenenbaum (2009) proposed that our ability to make strong causal inferences from small amounts of evidence must be in virtue of our having strong preexisting structured prior representations. They propose people represent their knowledge about the world in a hierarchy of abstract to specific domain knowledge about the plausible relations between the classes and types entities. When observing familiar causal relata they can then put more weight on the structures and functional forms that match their past experiences (Kemp et al., 2010). For example, infants as young as 5 months already possess an intuitive understanding of the laws of physics which leads them to be surprised by ostensibly implausible physical phenomena, such as objects seeming to jump appear or disappear rather than moving smoothly in continuous space (Newcombe et al., 2005). Moreover, Schulz and Gopnik (2004) have shown how

4 year-old children use domain knowledge (i.e. biology, physics or psychology) to support inference about particular causal relationships and are able to craft informative interventions to test their theories. Further work has proposed that domain-specific causal knowledge forms the backbone of children's ability for counterfactual reasoning (Sobel et al., 2004, 2007), since this inherently involves generating alternative states of affairs via one's pre-existing causal beliefs. They described domain specific priors as expert knowledge, providing insights into potential mediators, i.e. latent causal pathways, about co-occurring events, allowing the reasoners to make powerful extrapolations from sparse data.

While the Bayesian modelling implies people should represent their beliefs about the world as hierarchies of connected probability distributions, this is generally intractable to do explicitly.Various research has tried to reconcile this disconnection. Some studies have shown that people are conservative in their updates (Phillips and Edwards, 1966) while other evidence suggests that in some cases, participants fail to consider appropriate prior knowledge (Tversky and Kahneman, 1974; Stengård et al., 2022). Recent work has attempted to reconcile normative theories with suboptimal behaviours by drawing on the interpretation of such errors as resulting from an approximate inference process (Zhu et al., 2020; Dasgupta et al., 2020). We note that while expert knowledge can be meaningfully attributed to *prior* information brought in for the purpose of a new situation, interpreting deviations from a normative account in belief updating is challenging as it can be understood symmetrically as over-valuing the prior or under-valuing evidence (Jones and Love, 2011). As such, while a meaningful question to tease apart in continuous time, the present study will focus on the interpretation of priors as domain-specific knowledge.

Previous research on causal structure learning in continuous time focused on the algorithmic complexity of learning. It focused on identifying the key parameters of inference models which best match participants belief updating *within* a task rather than on interactions with the prior knowledge participants bring *to* the task (Bramley et al., 2017a,b; Davis et al., 2020; Gong et al., 2023). Uncontextualised experimental scenarios, while provided cleaner settings for formal analysis, miss out on a key dimension of causal learning, namely the integration of new data with the presumably rich pre-existing causal beliefs held by any adult reasoners. This paper take a first step into studying such interactions in the continuous-time active learning context.

1.4 Ornstein-Uhlenbeck networks

Ornstein-Ulhenbeck (OU) networks have recently been developed as a task environment and modelling framework for the study of continuous time causal structure learning Davis et al. (2020). An OU network is a collection of interdependent continuous Gauss-Markov processes, which results in a mathematically convenient model class for representing causal influences among continuous variables subject to noise, simulating causal dynamics and performing inference from dynamics to structure. OU networks can be represented as dynamical Bayesian networks, where the variables at a given time t are causes of themselves at time $t' > t$ (see Figure 1 for example paths). This allows for the representation of more complex causal structures involving feedback loops. The Ornstein-Uhlenbeck (OU) process (Uhlenbeck and Ornstein, 1930) is widely used to model naturally occurring phenomena. In physics, the Wiener

process, the simplest form of the OU process, otherwise known as Brownian motion, is used to model the movement of particles (Kramers, 1940) or the behaviour of a stock over a short time frame (Mandelbrot and Taylor, 1967). These applications provide a basis to use networks of OU processes to model causal dynamics among continuous variables. While a similar formulation has been used to model causal induction in simple tasks (Kruschke, 2006; Gershman, 2017), little research has used them to study how people reason about complex continuous time and state space systems.

The specifics of our adapted framework are described in the Formalism 3.1 section below. We also precisely describe our measure for accuracy, used throughout the following work, in Supplementary Materials 4. For another detailed account, see Davis et al. (2020).

2 Overview of experiments

All four experiments will be conducted in the same general framework. The principle, borrowed from Davis et al. (2020), is to provide participants with an interface with three sliders representing three variables and a graph plotting in real time the values of those variables (Figure 3.B.). The values of the variables are computed according to Ornstein-Uhlenbeck processes and depend on one another according to a latent causal graph structure. The participants' task is to observe the variations while also controlling the sliders to intervene on the variables, and thereby attempt to recover the latent causal structure. We now explain the two key interfaces which fully describe the paradigm. A demo for the interface can be found here Additionally, data and analyses scripts can be found here.

2.1 Common procedure

First, participants followed an illustrated introduction to causality by using informal graphical models. The full introduction can be found in Supplementary Materials 1.1. We introduced them to causal modelling by showing them a toy example of how to represent causal relations between a pandemic, a financial crisis and job losses. The causal strength of links was expressed simply by showing that given two nodes A and B, one could draw one arrow to indicate a moderate causal relation and two arrows to indicate a strong causal relation. The direction of the relationship was visualized with negative or positive sign next to the arrow. Participants were told that positive relationship implied that that higher values of the cause variable led to increases in the effect variable. Alternatively, negative relationships implied that higher values of the cause led to decreases in the effect variable. To preserve the relative distance between links, all judgements about graphs were collected using sliders with discrete values in {-2: strong inverse, -1, 0: no connection, 1, 2: strong direct}. We then showed participants illustrations of the three main building blocks of causal models: chains, common cause and collider. Following this, participants were presented with three graphs each with three variables. For these two graphs, and throughout the rest of the experiment, reporting causal links was always done using the scales shown on the right of Figure 3 A. The first graph only had one link which they simply had to report on the right hand side. The second had three links and was a common cause with

an additional link going from one of the effects to the other. The last one was a small scenario where variables were called flu, cough and asthma and participants had to report the causal graph induced by the following short text: "Tom has a cough. The doctor thinks that it could be a symptom of either asthma or the flu. He also believes that the flu is twice as likely as asthma to be the cause of the cough. The doctor also read a research paper stating that having asthma increased the chances of catching the flu". They could only proceed after correctly recovering the causal graph induced by the text.

Once participants were familiar with this representation of causal relations, priors were elicited using the interface presented in Figure 3 B. In the *criminality* domain, participants were first asked to picture themselves as the mayor of a city which had an issue with crime and was trying to understand the complex interplay between *crime*, *police* and *population happiness* (Supplementary Materials 1.2). They were then presented with an interface to report their intuitive estimate of each link values in the causal graph (Figure 3.A.).

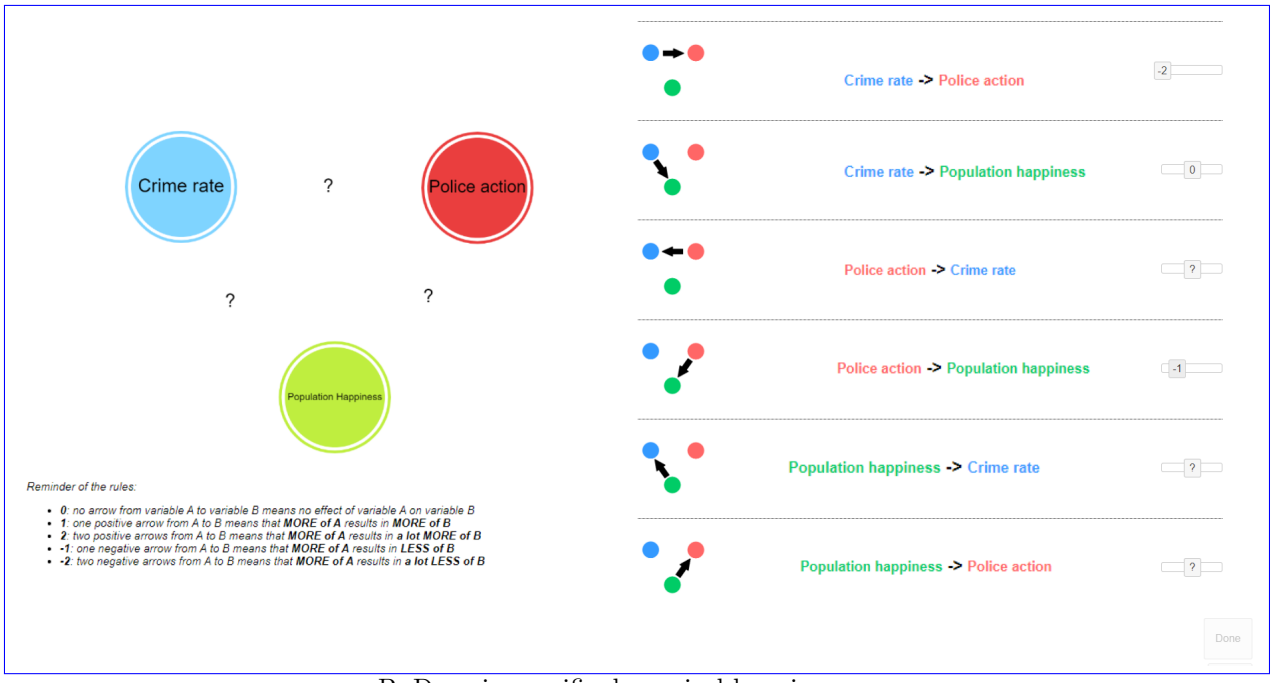
The second main interface was the dynamical causal learning game (Figure 3.B.). We showed participants an explanation of the rules of the game and a 40 second video showing an example of how to intervene and provide causal judgement. To proceed further, they had to provide correct answers to a multiple-choice comprehension test. In the game, we let participants intervene on each variable using simple sliders. Only one intervention at a time is allowed but they can last for as long as they wish and take, at any point, any value in the range $[-100, 100]$. While the range of values was displayed on the plot (see left side of Figure 3.B.), we did not provide participants with the actual values of the variables at each time ~~points~~ point so that participants did not attempt to make explicit calculations. Given that we let participants build their own graphs and that link strengths could be weak or strong, we provided visual aid in order to cope with unpredictably complex dynamics in the form of a graphical interface which dynamically plotted the last 10 seconds of data points generated by the OU network (Figure 3.B.). For consistency, the handles' vertical positions followed the last value on the graph. Interventions were plotted with a shading of the colour of the intervened variable. In the generic context (without any domain-specific labels), variable were named by their colours, i.e. *blue*, *green* and *red*. Thus a blue shading at some data points on the graph indicates that these were generated under an intervention on the *blue* variable. To summarise, the main trial interface was designed with the following constraints:

- Information: trials lasted 60 sec at a update rate of 5 Hz, i.e. $dt = 1/5$, which generated 300 data points in each trials.
- Visual aid: participants were presented with a plot of showing 10 seconds of history, interventions were plotted as background shading of the colour of the relevant variable.
- Memory: we allowed participants to report links during the trial and for an unlimited amount of time after, while still showing the last 10 sec of the trial on the plot. The goal was to avoid forcing participants to remember the links for a whole minute before reporting them using a newly learned interface.
- OU network parameters: $\theta = 0.5$ which describes the general strength of all causal effects in the graph, it multiplies into all causal relations and does not represent single causal links. In general, higher values of θ

will mean that variables in the network are more responsive to their causes. $\Gamma_{ab} \in \{-1, -0.5, 0, 0.5, 1\}$ which represents the causal link going from variable a to variable b and can be positive and negative and either weak, i.e. 0.5, or strong, i.e. 1. $\sigma = 3$ which is the standard deviation of the processes over one second, $dt = 1/5 = 0.2$ which is the time step, represents the inverse of the number of frames per second and scales the two previous parameters. We discuss those in details in Section 3.1.

We know from Rehder et al. (2022) that higher update rate dt and lower θ can lead to significant differences in performance by rendering causal effects respectively more sluggish or less impactful. Therefore, we adopt a form where the update rate $dt = 200$ms which is within the ranges they tested and a $\theta = 0.5$, which because our parametrisation is ever slightly different, is equivalent to their $\omega = \theta dt = 0.1$, which is the value they used in their rigid, i.e. more responsive, condition. All parameters were kept constant in all three experiments.

A. Prior elicitation interface



B. Domain specific dynamical learning game

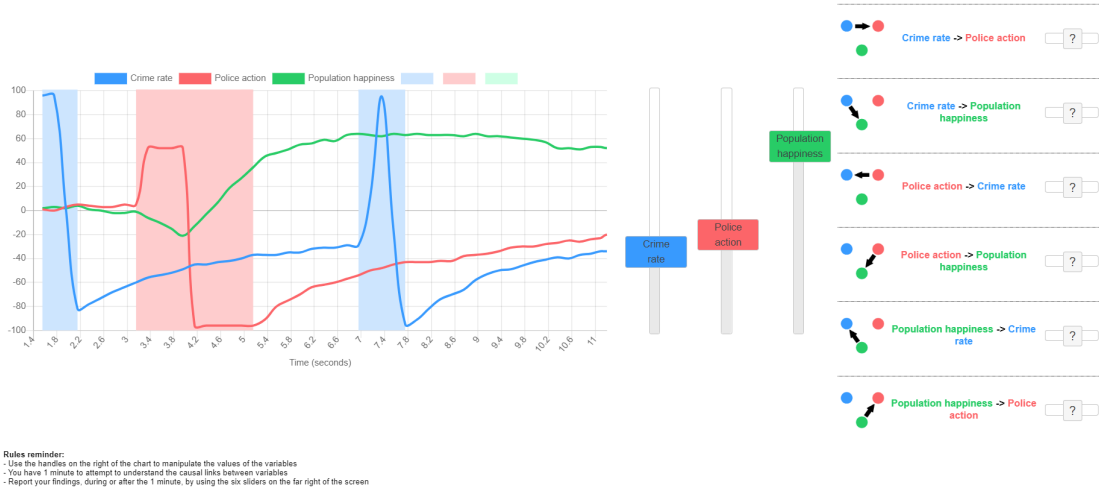


Figure 3: Prior elicitation and domain-specific interface for the *criminality* scenario in experiments 1, 2 and 3

2.2 Hypotheses

Given previous findings in active causal structure learning and domain-specific priors in causal reasoning, we detail below our main hypotheses.

1. We expect to replicate the predictions of the local computations theory of causal learning, namely that people will make more errors when considering indirect causal links in their final judgements.

2. We expect participants' judgements to be best captured by computational models that selectively consider data

during interventions and which focus on the direct links from the intervened upon node.

3. We expect participants to be less accurate in domain-specific contexts when the ground truth graph diverges from their prior judgements.

2.3 Experiment 1

In our first exploratory experiment, we wanted first to test whether participants, given a fixed ground truth causal structure, would perform differently if the variables were labelled meaningfully, i.e. *crime rate*, *police action* and *population happiness*, or generically, i.e. *blue*, *red* and *green*. Second, we wanted to see if the agreement between elicited prior and ground truth graphs would predict participants' ability to recover the correct structure.

2.3.1 Participants

60 participants (female 24, age 26.3 ± 6.6) were recruited via Prolific. Participants were paid on average £7.41 per hour, fees were fixed and independent from performance. The duration of the experiment was 26.8 ± 11.3 minutes. The present and following studies were approved by our institution's ethics committee (EP/2017/005).

2.3.2 Design

Participants were randomly allocated to a *generic* or a *informative label* condition. In the *informative label* condition we elicited participants' prior beliefs about the interplay between crime rate, police action and population happiness in criminal context. After three generic dynamical causal learning trials, in the fourth trial the variables were labelled to correspond to the criminal context. The ground truth of the fourth trial was kept fixed irrespective of participants' priors (Figure 4).

For participants in the *generic* condition, we did not elicit priors and the labels of the variables in the fourth trial were generic: blue, red, green. The ground truth was the same as in the *informative* condition. For all other trials, both conditions were identical. Due to some participants not completing the task, the split was not perfect (27: *generic*, 33: *informative*).

2.3.3 Procedure

After providing informed consent, participants went through a basic introduction to causal models to familiarise them with the graph representation of links and variables. They were then asked to learn about the interface we used to report links by representing models we showed them using it (see Figure 3). It was explained that connections could be marked as either negative, using a "$-$", or positive, using a "$+$". Moreover, the strength of an effect could either be normal, visualised as one arrow, or strong, visualised as two arrows. The same reporting interface was used in the main task. Participants had to report on three models and could not move on before correctly reporting them.

For participants assigned to the *informative* condition, we elicited prior guess about the causal structure in the same way we later elicited their evidence-based structure judgments (see Figure 3 A. and B.). Participants assigned

to the *generic* condition simply skipped this step.

Dynamical structure learning task In the second section of the experiment, participants watched a brief video presentation on the task and the interface, completed a comprehension quiz and then completed 4 test trials. In the first three trials the true structure was, in random order: causal chain, common cause and common effect, for which all links were strong. For each of these, we had two variants, one where all links were positive and the other where all links were negative. The variants were counterbalanced between participants. The fourth was the graph we call damped feedback loop and defined to be a possible representation of the interplay between Police action, Crime Rate and Population Happiness resulting in a trial depicted in Figure 3. Participants assigned to the *generic* condition performed the same trial but labels were: blue, green, red (Figure 4).

A. Generic condition Damped Feedback Loop B. Label condition Damped Feedback Loop

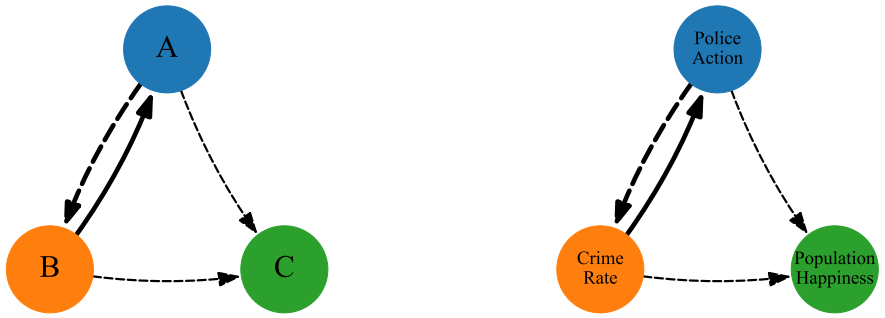


Figure 4: Damped feedback loop graphs used for the last trial of experiment 1 in the generic condition (A.) and in the informative condition (B.). They both share the same structure. Full lines indicate positive links and dashed line negative links. Wider arrows indicate strong links and thinner arrow weak links.

2.3.4 Results

Table 1: Frequencies of recovered link and accuracy in model recovery for all trials

Graph	Individual links recovered[1]	Graph accuracy[2]
Mean recovered	$.627^{***}$	$.829^{***}$
Damped Feedback Loop (Generic)	$.558^{***}$	$.808^{***}$
Damped Feedback Loop (Label)	$.549^{***}$	$.799^{***}$
Chain	$.589^{***}$	$.821^{***}$
Negative Chain	$.568^{***}$	$.767^{***}$
Collider	$.703^{***}$	$.865^{***}$
Common Cause	$.678^{***}$	$.856^{***}$

*** $p < .001$ $N = 60$ [1] One sampled t-test against the chance level of 0.2
[2] One sampled t-test against the chance level of 0.4. Note that this is not a probability but a measure of the average distance from the ground truth (see Supplementary Materials 4).

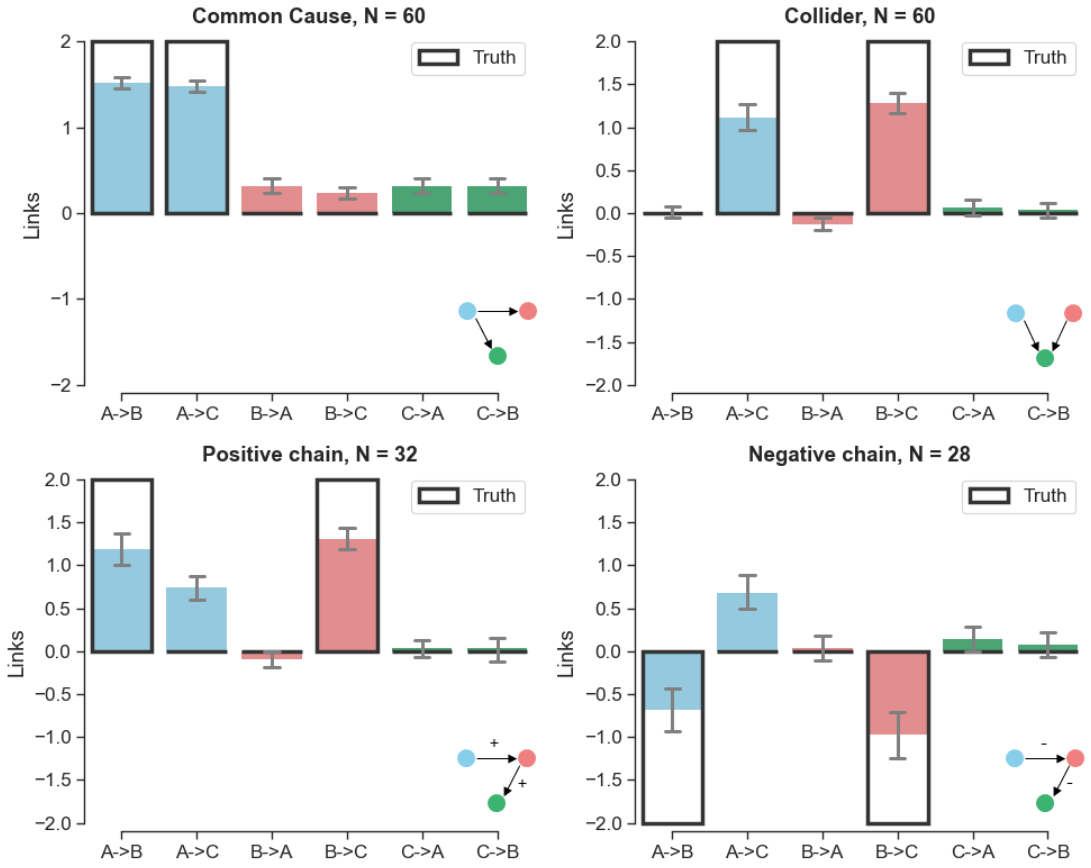
In accordance with previous research, we find that participants were overall very good at reporting correct links,

i.e. reporting the exact link value, even considering the fact that they tended to confuse strong and weak links. Compared to the chance level of .2, they recovered correct link values .63 of the time on average ($\mu = .63, t(59) = 15, p < .001$). Overall, compared to eyes-closed chance accuracy (.33 as links can either be positive, negative or null), participants recovered .87 of positive links ($\mu = .87, t(59) = 18.08, p < .001$) while they recovered .73 of negative links ($\mu = .73, t(59) = 10.40, p < .001$). The difference in accuracy between positive and negative links was significant ($t(118) = 2.80, p < .01$). We expect that this difference is mainly due to the negative link definition ('more of A means less of B') being less intuitive than the positive definition ('more of A means more of B'). Participants performed significantly better than chance level (.33) at recognising weak and strong links but not by a strong margin. On average, they recovered ~~.4~~ .40 of weak links ($\mu = .4, t(59) = 2.69, p < .01$) and .50 of strong links ($\mu = .50, t(59) = 5., p < .001$). The difference between ~~both~~ weak and strong was not significant. These results, compared with how effective participants' were at recovering the polarity of the connections, suggests that they struggled most to distinguish between weak and strong links. This was expected as they were not given visual training on the difference between a weak and a strong link before the task. This can be seen in Figure 5 where the average of reported links for strong links falls between 1 and 2. Participants made extensive use of intervention. If we count any period in which a slider is under control as a separate intervention, the average number of interventions per trial was 7.18 and each intervention lasted on average 3.38 ($\pm$3.96) seconds. The average proportion of each trial spent intervening was .44 ($\pm$.20)

We measure using a normalised measure of ~~euclidean~~ Euclidean distance between ground truth and posterior judgement graph where 1 equates to reporting the exact ground truth and 0 equates to reporting the furthest possible graph from the ground truth (see Supplementary Materials 4 for a detailed explanation). Under this measure, participants were far above chance at approximating chain and damped feedback loop graphs, with respective mean accuracies of $.79 \pm .11$ and $.80 \pm .13$ (see Table 1 column 3 for details about variants of each graph). Performance was better for simpler graphs, i.e. collider and common cause, where accuracies were respectively, $.87 \pm .24$ and $.86 \pm .24$. These findings are in line with the theoretical predictions of the LC model (Fernbach and Sloman, 2009) as we can see that participants both reported a non-existent direct link in the chain graphs (Figure 5 A., links $A \rightarrow C$ in positive and negative chains) and failed to report a damped direct link in the damped feedback loop graph (Figure 5 B. link $Police \rightarrow Happiness$ and $A \rightarrow C$). In the latter case, despite $Police$ being a direct negative cause $Happiness$, the indirect action of $Police$ on $Happiness$ through $Crime$ is positive, yielding an ambiguous situation where an increase in $Crime$ will cause an immediate decrease in $Happiness$ but which will quickly disappear due to the counteracting indirect effect through $Crime$. While the pattern in the data between this structure and a case where $Crime$ does not cause $Happiness$ are subtly different, a prediction of the LC model is that participants will tend to favour the one where no link is present as they will fail to consider the possibility of indirect effects. In addition, we see that the graph we proposed for criminality was understandable with a complexity lying in the same range as a chain model.

Average accuracy in the *generic* condition was not significantly higher ($.81 \pm .13$) than in the the *informative* condition ($80. \pm .13$). We found no evidence that accuracy depends on whether the variables were informatively

A. Generic graphs



B. Damped Feedback Loop graph in control (left) and label (right) conditions

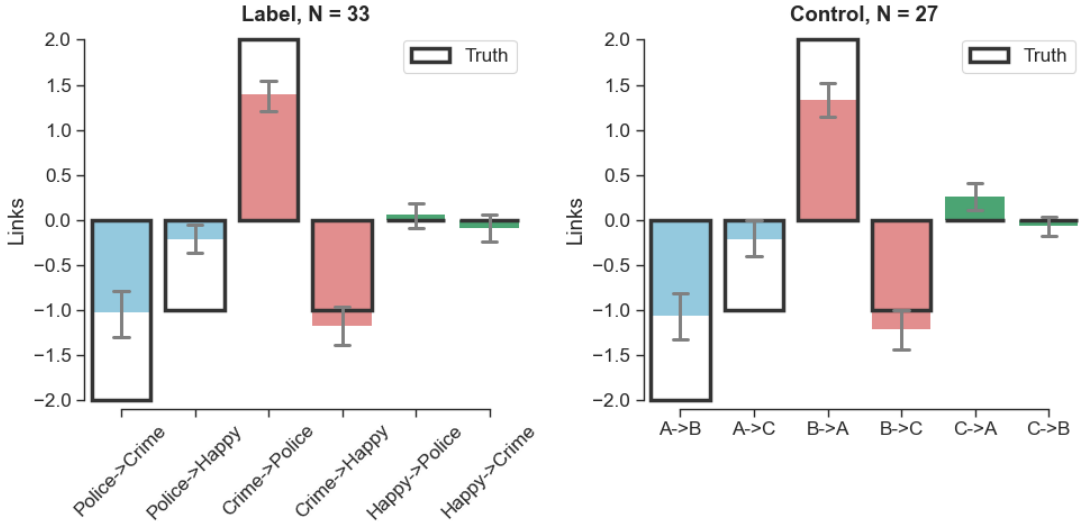


Figure 5: Graphs generated by participants in all trials, colours represent judgements and black square ground truths. The x-axis represents the link and the y-axis its strength (2: strong positive link, 1: weak positive link, 0: no relation, -1: weak negative leak, -2: strong negative link)

labelled. The difference between conditions was not significant (~~$t = -.26, p = .79$~~$t(58) = -.26, p = .79$). To investigate whether participants' posterior judgements in the label condition were influenced by their prior, for each of the 33 participants for whom we have collected priors for the damped feedback loop, we predict their accuracy in the last trial from the proximity between the ground truth graph and their prior (*Prior GT proximity* $\in [0, 1]$). We expect participants with closer priors to be more accurate. ~~We also compute the~~ To control for participants' performance in the task, i.e. irrespective of the causal structure, we add their average accuracy across generic trials as a measure of overall ability ~~in the task~~ (*mean accuracy*). We fit a linear model predicting accuracy in the criminality trial from these two quantities. Both *prior GT proximity* ($\beta = .44, t(31) = 3.989, p < .001$) and *mean accuracy* ($\beta = .67, t(31) = 4.817, p < .001$) were highly predictive of *damped feedback loop accuracy* ($R^2 = .65, F(2, 31) = 28.8, p < .001$). That is, participants whose prior was closer to the ground truth were on average more accurate.

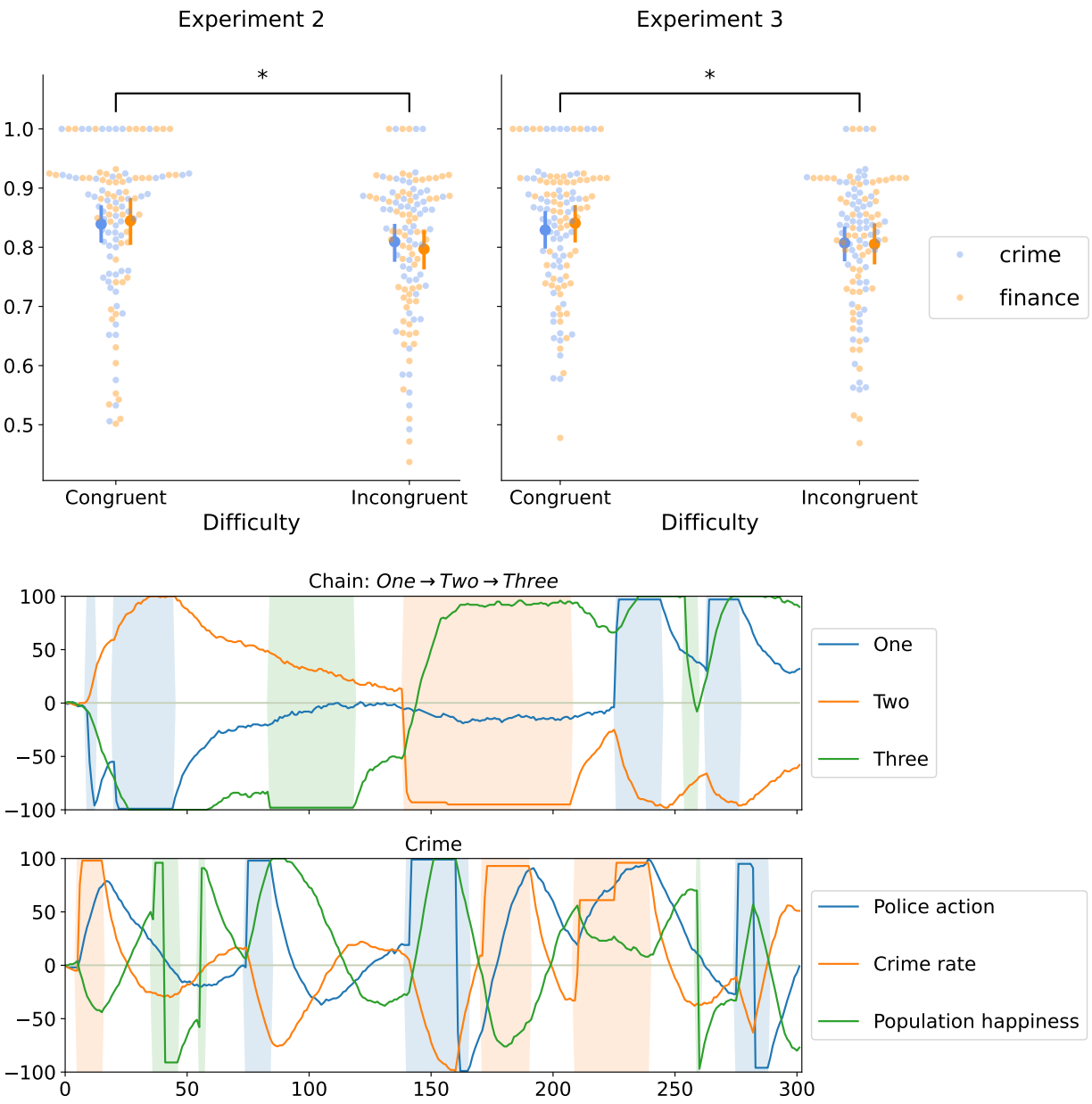


Figure 6: Congruence and scenario group means and swarm plots for accuracy. Brackets indicate pairwise t-tests results. Sample paths for two participants, one through a chain trial in experiment 3 and crime trial in experiment 3.

2.4 Experiment 2: reversed connections

One shortcoming of Experiment 1 was that the relationship between ground truth for the test trial and participants prior beliefs was not controlled.The correlation between the agreement of participants' prior judgements with the ground truth and their final judgment accuracy in Experiment 1 is suggestive of a supporting influence of real-world expectations but testing this thoroughly requires a more controlled experiment. To investigate this effect, we used a

design where we dynamically manipulated the ground truth depending on participants' prior judgements.

We elicited priors across two scenarios (*criminality* and *finance*) and for each scenario manipulated the ground truth graph to make the match with their prior either good, *congruent*, or bad, *incongruent*. This study addresses some confound stemming from the design in ~~experiment 2 which has led us to exclude it from further analyses (~~an exluded experiment, the details and results of which are presented in Supplementary Materials 2). The graph in the two conditions were defined as follows: *congruent*, where the ground truth graph was exactly the prior and *incongruent*, where the direction of all the links in the priors was reversed, i.e. $A \rightarrow B$ became $B \rightarrow A$ and so on. This method controls for the density of the causal model while making the prior either helpful or a presumed hindrance.

2.4.1 Participants

120 participants (female: 35, age 28.9 ± 9.5) were recruited via Prolific. Participants were paid on average £6.01 per hour, fees were fixed and independent from performance. The duration of the experiment was 31.41 ± 13.42 minutes. This study's desired sample size, included variables, hypotheses, and planned analyses were preregistered on Open Science Framework (link removed for anonymity) prior to any data being collected.

2.4.2 Design

We defined the new *congruent* condition such that the ground truth was exactly the prior judgement and the new *incongruent* condition such that the ground truth mirrors prior judgement graph but the direction of all links has been reversed, e.g. a chain $A \rightarrow B, B \rightarrow C$ is now $C \rightarrow B, B \rightarrow A$. While this leads to a qualitatively different causal model in many instances, e.g. a common cause becomes a collider, this transformation preserves the complexity of the structure in terms of how many edges it contains and preserves loops and chain structures which are the more profound sources of confusion for participants. With now two difficulties, we add a new scenario we call *Finance*. The variables for the *Criminality* scenario were Crime rate, Police action and Population happiness. The variables for *Finance* were Virus cases, Stock prices and Lockdown measures.

The resulting design is a within-subject incomplete factorial study with $2^2 = 4$ congruence-scenario pairings, of which each participant sees only 2. The order and congruence pairing of the scenarios were randomised.

2.4.3 Procedure

Causal graphs introduction This section was identical to Experiment 1.

Dynamical structure learning task Participants completed four trials. The first trial was a randomly selected generic graph to train participants with the interface. The ground truth was either a chain or a chain in which a direct link was added between the first and last variable. The position of all variables was randomised and links could be negative or positive.

After the first training trial, participants completed two labelled trials, separated by another generic trial. The labelled trials were randomly selected from the pool of four congruence-scenario pairs such that each participant sees exactly each scenario and each congruence condition once. Priors were elicited using the method presented in section 1.2 for all labelled trials.

2.4.4 Results

As in Experiment 1, participants were proficient in recovering correct links, reporting .59 of correct links overall, performing much better than the .2 chance level ($\mu = .59, t(119) = 27.24, p < .001$). Participants also were better at recovering positive ($\mu = .77, t(119) = 21.13, p < .001$) than negative ($\mu = .67, t(119) = 14.95, p < .001$) links ($t(238) = 3.14, p < .01$). In a pattern consistent with previous experiments where participants struggle with link strength, they recovered weak links ($\mu = .42, t = 1.26, p < .21$) not better than chance level (0.33) and significantly above it for strong links ($\mu = .41, t = 3.24, p < .01$). The difference was not significant. Given this, we ran an analysis where we collapsed links over strength, leaving only negative, no link and positive, which did not change the overall pattern of results (see Supplementary Materials 7). In this light and as participants used weak and strong link to represent their priors, we chose not to collapse links over strength. We found no order effect of either scenario or congruence, suggesting that the order in which participants were exposed to the conditions did not affect their performance. As in experiment 1, participants made plentiful use of interventions. The average number of interventions per trial was 8.41 and they lasted on average 3.02 ($\pm$3.37) seconds. The average proportion of intervention time with respect to the total available time was .44 ($\pm$.19).

We fit a linear mixed effect model predicting *accuracy* from *congruence* and *scenario* and their interaction with random intercepts at the participant level. The results are presented in Figure 6. There was no significant fixed effect of scenario nor significant interaction between scenario and congruence. There was a significant fixed effect of congruence on accuracy ($F(122.56) = 6.22, p = .014$). Indeed, participants were on average less accurate by .039$\pm$.016 in the *incongruent* condition than in the *congruent* condition ($t = 2.494, p = .014$), suggesting priors judgements indeed affect posterior estimates. We note again that participants did remarkably well in all circumstances and that while significant, the effect was quantitatively small (.039$\pm$.016 on a $[0, 1]$ scale). This suggests that while participants were clearly able to map their abstract prior knowledge onto dynamics unfolding in continuous time, this process led to only a small difference in performance.

2.5 Experiment 3: conservative replication of ~~study 3~~Experiment 2

In our final experiment, our aim was to replicate the results of Experiment 3 but with a design in which the prior elicitation was more temporally split from its associated dynamical learning game trial. We moved the two prior elicitation phases to the start of the experiment, and separated them from the informative label trials with two generic label trials. By adding a gap between these steps, our goal was to control for potential effects stemming from playing the causal learning game immediately after reasoning about the labelled variables in the prior elicitation trial. The

aim was to replicate the findings of the study 3 with a more conservative design and observe whether the effect of prior elicitation still occurred.

2.5.1 Participants

120 participants (female: 56, age 27.6±9.3) were recruited via Prolific. Two were removed, due to a data saving issue. Participants were paid on average £6.81 per hour. Fees were fixed and independent from performance. The duration of the experiment was 32.00 ± 13.38 minutes. This study's desired sample size, included variables, hypotheses, and planned analyses were preregistered on Open Science Framework (link removed for anonymity) prior to any data being collected.

2.5.2 Design

The design was identical to study 3 except that we split the prior elicitation phases from the associated labelled dynamical structure learning trials by having two generic dynamical trials in between.

2.5.3 Procedure

Causal graphs introduction This section was identical to experiment 1.

Prior elicitation We elicited priors using the previously introduced procedure for both scenarios

Dynamical structure learning task Participants first saw 2 generic trials, 1 labelled trial, 1 generic trial and finally the second labelled trial.

2.5.4 Results

As in the previous studies, participants recovered .63 of links ($\mu = .63, t(117) = 23.78, p < .001$). They did better with positive links ($\mu = .80, t(117) = 20.24, p < .001$)) than with negative links ($\mu = .69, t(117) = 14.90, p < .001$); the difference was significant ($t(234) = 3.27, p < .01$). Participants still did better than chance for weak ($\mu = .42, t = 3.81, p < .001$) and strong ($\mu = .46, t = 4.30, p < .001$) links. The average number of interventions per trial was 9.83 and they lasted on average 3.07 (±3.23) seconds. The average proportion of intervention time with respect to the total available time was .42 (±.19).

We perform the same main behavioural analysis as experiment 2. We fit a linear mixed effect model predicting *accuracy* from *congruence* and *scenario*, letting intercepts vary per participants. In accordance with our preregistration, we add a further continuous variable, *Local Computation Agent Score*, which is the accuracy on each trial of a normative local computations agent and serves as a control variable to make sure that differences in accuracy stem from the effect of prior graphs and not from participants being faced with graphs that are structurally harder to learn in incongruent trials, i.e. graphs with more indirect effects and thus generating more ambiguous data for agents using a local computations strategy. These scores are generated by having a local computations model perform inference

on each graph while selecting optimal actions (see Appendix 6.1). Its final MAP estimates is used to compute its accuracy with respect to the ground truth. The results of the mixed effect model can be seen on Figure 6. We found a significant fixed effect of *congruence* on *accuracy* ($F(114.49) = 4.83, p = .03$), suggesting a robust effect whereby *accuracy* was boosted in trials in which the ground truth agreed with participants priors and hindered when the reverse was true. The difference between *congruent* and *incongruent* trials was on average .029 ($t = 2.20, p = .03$). There was no fixed effect of *scenario* nor any significant interaction. There was no fixed effect of *Local Computations Agent Score*, suggesting that the graphs were not structurally harder to learn in either conditions. The order in which scenarios were presented did not affect participants' accuracy. Participants who performed the incongruent trial before the congruent trial did significantly worse than the ones that were presented the trials in reverse order ($F(2, 116) = 7.44, p < .01$). We note that given the design, in the first case, the incongruent trial was the third out of five total trials while in the second case, the incongruent trials was the fifth. We therefore suppose that the detrimental effect of incongruence on accuracy is relatively short lived and that for the latter group, by the fifth trial it has almost fully dropped off. This set of studies does explore this phenomenon further but we posit that this may be an instance of cued recall whereby data congruent with prior knowledge may help participants recall their prior graph (Schooler and Anderson, 1997; Lohnas and Kahana, 2014). Given congruent dynamics, participants may more easily realise that the underlying graph is their prior. Alternatively, when the dynamics are incongruent, then the priors are not recalled unless they are still fresh in memory, hence explaining this order effect.

3 Modelling

In this section, we seek to provide a deeper understanding of participants' behaviour using computational modelling. This will allow us to formalize participants' inference process and the role of prior knowledge in labelled trials. Previous research found that a Causal Event Abstraction model, a heuristic model departing from Bayesian inference, best captured participants' judgements (Rehder et al., 2022). Here, we show that an alternative Bayesian account, which allows prior knowledge to influence subsequent inference, but only regards data collected during interventions, provides a better account of human continuous time causal learning.

3.1 Formalism

Ornstein-Uhlenbeck (OU) process The Ornstein-Uhlenbeck process (Uhlenbeck and Ornstein, 1930) is a continuous-time stochastic process which is related to the Wiener process (a continuous-time Gaussian-Markov process, or random walk), with the important difference that the processs reverts to the value of an attractor term μ with a force parametrised by θ. The probability that the process X takes value x at time $t + dt$ is given by the following Gaussian distribution:

$$p(x_{t+dt}|x_t, \mu, \theta, \sigma, dt) = \mathcal{N}(x_t + \theta(\mu - x_t)dt, \sigma dt) \tag{1}$$

where dt is the timestep, σ is the variance, μ is the attractor and pulls X towards its value and θ is a scalar parametrising the strength of this attraction as an increasing function of the distance between μ and x_t.

Ornstein-Uhlenbeck networks Davis et al. (2020) describe Ornstein-Uhlenbeck (OU) networks as a collection of OU processes, i.e. random variables in continuous time, for which the attractors μ are parametrised by relations in a causal network. If X and Y are two random variables described by OU processes, μ can be used to represent a causal link between both variables. As a function of the values of X and Y at time t, $\mu_x(x_t, y_t)$ is defined to be the value of the attractor of X conditional on the value y_t, one can see how it would make x_{t+dt} the value of X at $t + dt$ dependent on Y at time t:

$$p(x_{t+dt}|x_t, y_t, \theta, \sigma, dt) = \mathcal{N}\left(x_t + \theta(\mu_x(x_t, y_t) - x_t)dt, \sigma dt\right) \tag{2}$$

where $\mu_x(x_t, y_t) = x_t + \gamma_{yx} y_t$ is a linear function of its inputs and γ_{yx} represents the causal strength of x on y. As the context is causal, the order matters and γ_{yx} denotes the effect of Y on X. Notice that when $\gamma_{yx} = 0$, $\mu_x(x_t, y_t) = x_t$ and we have that x_{t+dt} does not depend on y_t. Alternatively, if γ_{yx} is negative then the causal effect is reversed, where higher positive values for y_t will lead to decreases in x_{t+dt}. It should be noted that Bayesian networks cannot encode the causal loops allowed by this equation. It is however possible to represent OU networks as dynamical Bayesian network where each variable at each time point is a different node with arrows going into it from all variables at the previous time point (Rehder, 2017).

In the present study, the range of values that variables can take is bounded $x_t, y_t \in [-100, 100]$, we add a $\beta(x_t) = -\gamma_{xx} x_t \frac{|x_t|}{100}$ regularisation term to μ_x which depends on x_t and moderates the attractive force of X back to the origin. The force is weak when close to the origin and strongest when close to the bounds (see Supplementary Materials 3 for a full explanation).

Causality Matrices Given the aforementioned parameters θ and σ and time step dt, an OU network for K variables $\mathbf{X} = \{A, B, \ldots\}$, is fully described by its causality matrix Γ, a $K \times K$ asymmetric matrix with entries Γ_{ij}. Each Γ_{ij} is the causal effect of the variable at index i, say A, on the variable at index j, say B, or equivalently $\Gamma_{ij} = \gamma_{ab}$. For simplicity, the causal dependence of each variable onto itself is always set to $\Gamma_{ii} = 1$. Additionally, Γ can be represented a $K^2 - K$ vector $\boldsymbol{\gamma}$ composed of all its non-diagonal entries.

In this set of studies we focus on networks with 3 variables. In Davis et al. (2020), possible values for Γ_{ij}, i.e. γ_{ij}, were restricted to the discrete set $\{-1, 0, 1\}$, allowing positive and negative links. In contrast, we extend the set to include weaker links. Thus in all cases, the values for Γ_{ij} were in $\{-1, -0.5, 0, 0.5, 1\}$. For networks with 3 variables, i.e. $\mathbf{X} = \{A, B, C\}$, Γ is 3×3 and $\boldsymbol{\gamma}$ is a $K^2 - K = 6$ dimensional vector. Γ is given by:

$$\begin{pmatrix} A \to A & A \to B & A \to C \\ B \to A & B \to B & B \to C \\ C \to A & C \to B & C \to C \end{pmatrix} \implies \Gamma = \begin{pmatrix} 1 & \gamma_{ab} & \gamma_{ac} \\ \gamma_{ba} & 1 & \gamma_{bc} \\ \gamma_{ca} & \gamma_{cb} & 1 \end{pmatrix}$$

Notice that this represents the adjacency matrix of a 3 variable directed graph and the space of all possible Γ represents a space of graphs. We consider these to be causal graphs but not in the formal sense of causal graphical models (Pearl, 1997) as loops are allowed. Moreover, as each non-diagonal entry can take a value in $\{-1, -0.5, 0, 0.5, 1\}$, the full sample space contains $5^{K^2-K} = 15625$ graphs.

Finally, this makes updates very simple to represent in matrix form. Let $\mathbf{x}_t$ be the vector of values of all variables in the network at time t, the transition probabilities for all variables in the network becomes a multivariate Gaussian given by:

$$p(\mathbf{x}_{t+dt}|\mathbf{x}_t, \Gamma, \theta, \sigma, dt) = \mathcal{N}\big(\mathbf{x}_t + \theta(\boldsymbol{\mu}_t(\mathbf{x}_t) - \mathbf{x}_t)dt, \sigma dt I_K\big) \tag{3}$$

where $\boldsymbol{\mu}_t(\mathbf{x}_t) = \beta(\mathbf{x}_t) + \mathbf{x}_t^\top \Gamma$ and I_K the K dimensional identity matrix. Note that θ affects all causal links to the same extent, when it is 0, all causal effects vanish and the network reduces to a set of uncorrelated random walks. This standard form should remind the reader of the assumptions about dynamics made by the widely studied linear Kalman Filter (Speekenbrink and Shanks, 2010; Gershman, 2015; Kruschke, 2006).

Interventions Finally, we follow Davis et al. (2020) by including interventions on the variables. Only one intervention is allowed at each time point. It acts as Pearl's *do* operator in that the value of any variable on which the participants intervene ignores its causal determinations and is set to the chosen value with probability 1 (Pearl, 2009). From an inference standpoint, we denote a_t the intervention at time t. a_t acts as an indicator such that the transition probability, from equation 2 becomes:

$$p(x_{t+dt}|x_t, y_t, a_t, \theta, \sigma, dt) = p(x_{t+dt}|x_t, y_t, \theta, \sigma, dt)^{\mathbb{1}[a_t \neq X]} \tag{4}$$

where $\mathbb{1}[a_t \neq X]$ is an indicator function returning 0 if $a_t = X$. In this case the left hand side will be equal to 1, essentially preventing updates on links going into a variable currently being intervened on. a_t can either be set to the variable itself, i.e. $X, Y, ...$, or equivalently to its index in the network $i \in \{1, 2, ..., K\}$.

3.2 Learning in Ornstein-Ulhenbeck networks

If we formally summarise learning as finding a posterior distribution over the whole space of possible graphs (Griffiths and Tenenbaum, 2005), we must consider that the agent hold prior distribution over all parameters, that is to say over θ, σ, dt and the possible matrix Γ induced by any possible graph $\mathcal{G}_K$, where K denotes the number of variables in the network. Ignoring the timestep dt, the full learning problem can by summarised in the following equation:

$$p(\Gamma|\mathbf{x}_{1:T}) \propto \prod_{t=1}^{T}\prod_{i=1}^{K}\int_\theta\int_\sigma p(x_t^i|\mathbf{x}_{t-1}, a_t, \Gamma, \theta, \sigma)p(\Gamma, \theta, \sigma)d\theta d\sigma \tag{5}$$

We note that the choice made by Davis et al. (2020) and ourselves to assume that participants estimate parameters which are not entries of Γ from the training phase is necessary to make learning straightforward as we can make no further assumption about the factorisation of $p(\Gamma, \theta, \sigma)$. Indeed, those integrals are not analytically solvable and

the full learning problem would require using approximate Bayesian methods such as Markov Chain Monte Carlo or variational inference (Sanborn, 2017). An initial attempt using mean-field approximation, i.e. assuming that $p(\Gamma, \theta, \sigma) = p(\Gamma)p(\theta)p(\sigma)$ and estimating posteriors for each parameter, has been proposed by Btesh et al. (2023a) but presently, the full inference process is thus simplified as follows:

$$p(\Gamma|\mathbf{x}_{1:T}) \propto p(\Gamma) \prod_{t=1}^{T} \prod_{i=1}^{K} p(x_t^i|\mathbf{x}_{t-1}, a_t, \Gamma, \theta, \sigma) \tag{6}$$

3.3 Modelling the influence of prior judgements

To model the influence of priors on posterior judgements, we start by focusing on a standard Bayesian account, i.e. that a prior judgement parametrises an initial distribution over graphs. Previous work (Davis et al., 2018, 2020; Rehder et al., 2022) has implicitly treated the prior distribution as uniform, that is $p(g \in \mathcal{G}) \propto 1$, where g is a causal graph and $\mathcal{G}$ the set of all possible causal graphs. We propose, from a prior judgement g_p, a distribution over graphs $p(g)$ such that the maximum weight is given to the prior judgement. To smooth out this prior so that graphs close to the prior judgement are given more weight that those further away, the distribution is computed by a softmax function with an inverse temperature τ_p and a vector of values for each $g \in \mathcal{G}$ which describes how distant each of them is from g_p. The prior probability of some graph g was thus given by:

$$p(g) \propto \exp\left(\tau_p \text{prox}(g, g_p)\right) \tag{7}$$

where $\text{prox}(g, g_p) \in [0, 1]$ is the proximity measure between graph g and the prior judgement graph g_p based on their Euclidean distance (see Supplementary Materials 4 for a derivation of this measure). $\text{prox}(g, g_p) = 1$ if and only if $g = g_p$ and will be lower otherwise, hence giving a higher prior probability to graphs which are closer to the prior judgement g_p. We note here that such a formulation of priors can represent a variety of potential mechanisms and thus should be considered a descriptive account. We will below consider how priors are used as expert knowledge, preventing errors while estimating indirect effects. However, this standard Bayesian formulation of priors can also represent more computational constraints, i.e. participants potentially dealing with the abundance and auto-correlation of datapoints in continuous time by devaluing the contribution of each new observation to their posterior. A standard Bayesian formulation such as the one in equation 7, will interpret such a phenomenon as equivalent to learning with a peaked prior. While we will later provide evidence for the expert knowledge interpretation, this work leaves for future research the task of disentangling and evaluating the contribution of both interpretations.

3.4 Competing inference models

We present computational models that have been previously used in this context with specific additions we believe to be relevant to capture participants' behaviour in domain specific contexts where links can be weak or strong. Recall that in general, the inference problem is formalised as learning the entries of a $K \times K$ causality matrix Γ, which is the adjacency matrix of a graph $g \in \mathcal{G}_\mathcal{K}$, where the non-diagonal entries are the values representing the strength of

each link in g. In all cases, we fit one set of parameter values to each participant. Furthermore, we do not attempt to model interventions, we only fit the models' parameters to participants' posterior judgements about the graph, i.e. the specific Γ, that generated the sequence.

3.4.1 Normative

We consider the case where the agent's model of the world aims to exactly match the world itself and the agent's task is simply to learn the parameters of the causality matrix Γ to infer the correct graph g^*. We consider that participants acquire estimates of the system's parameters: θ, which represents a general strength applied to all causal effects, σ representing the noise of the system, a time step dt, and the possible values of Γ by watching the presentation video and learning about possible link values in the causal graphical model introduction. Indeed, while participants do not see all possible link values in $\{-1, -0.5, 0, 0.5, 1\}$, they are explicitly told about the proportionality relationship that exists between weak and strong links, i.e. that the latter is twice as strong as the former. Given such estimates, the normative model solves the full simplified inference problem (equation 6). For any given link $i \to j$, where variables are represented by their indices and a vector of variable values $\mathbf{x}_t$, ignoring known parameters θ and σ, we get the following update rule:

$$p(\Gamma_{ij}|x^j_{t+dt}, \mathbf{x}_t, a_t) \propto p(x^j_{t+dt}|\mathbf{x}_t, \Gamma)^{\mathbb{1}[a_t \neq j]} p(\Gamma_{ij}) \tag{8}$$

where $\mathbb{1}[a_t \neq j]$ is an indicator function returning 0 if the agent is intervening on target variable j, i.e. $a_t = j$ and 1 otherwise. From the definition of a causal intervention, this will lead the model to consider the link $i \to j$ cut and to not update its belief about it. Notice how the update for a single link depends on the vector of all variables at the previous time point t and on the full causality matrix Γ. This follows from the transitions described by equation 3 and means that normative learning requires maintaining a belief distribution over all possible Γ. In networks with 3 variables and 5 possible link values, this yields 15625 hypotheses.

Parameters We fit the normative model with an inverse temperature parameter τ which controls how peaked the posterior distribution is on graphs, i.e. causality matrices Γ, with larger posterior probabilities. It is applied to the posterior after observing all datapoints. More formally, the final posterior of the model is given by the following softmax distribution:

$$p(\Gamma|\tau, \mathbf{x}_{1:T}) \propto \exp\left(\tau p(\Gamma|\mathbf{x}_{1:T})\right) \tag{9}$$

When τ approaches 0, $p(\Gamma|\tau, \mathbf{x}_{1:T})$ approaches a uniform distribution. Alternatively, when τ becomes large, $p(\Gamma|\tau, \mathbf{x}_{1:T})$ approaches a point mass on its maximum. The fitting procedure seeks the value for τ which maximises the likelihood of the participant's final judgement. Larger values of τ will signal that participants' judgements have higher posterior probability under this model.

To get a sense of the weight of priors in posterior estimates, we fit two versions of the normative model, one where we set the prior inverse temperature $\tau_p = 0$, equivalent to a uniform prior, and the other where we fit τ_p to participants' data, using the method described above (see Section 3.3). Conversely to the posterior inverse temperature, this one was applied before observing any data. This resulted in modelling starting each trial with a prior more or less peaked on each participant's prior judgements. This method was used for all models except for the Causal Event Abstraction model as it is not Bayesian.

3.4.2 Local computations

The local computations (LC) model follows the normative model, i.e. attends to all data points and holds estimates for all parameters, but importantly simplifies the causal dynamics by not holding a full distribution over Γ. Instead, it holds pairwise marginal distributions over each link, i.e. each non-diagonal entries of Γ. Essentially, it splits the inference over the full OU network into inferences over $K^2 - K = 6$ smaller two-variable networks. For instance, this means that when updating the link $A \to C$, it will only consider the values of A and C and ignore B. For any given link $i \to j$ in Γ, where i and j can be the indices of any unmatched pair (A, B), (B, C), etc., and ignoring known parameters, we get the following update rule:

$$p(\Gamma_{ij}|x^j_{t+dt}, \mathbf{x}_t, a_t) \propto p(x^j_{t+dt}, |x^i_t, x^j_t, \Gamma_{ij})^{\mathbb{1}[a_t \neq j]} p(\Gamma_{ij}) \tag{10}$$

where we have the same intervention indicator function as the normative model. Notice how compared to normative learning in equation 8, no variables other than i and j figure in this update. This is the fundamental assumption behind LC in this framework and it allows the agent to compress its representation of the hypothesis space from a distribution over 15625 graphs to 6 distributions over 5 values that links can take, i.e. one for each non diagonal entry in Γ.

Parameters The local computations model is fit with the same τ parameter as the normative model and also with a version with and without prior influence (τ_p).

3.4.3 Causal event abstraction

The causal event abstraction (CEA) model is a heuristic model proposed by Rehder et al. (2022) to show that a simple discrete heuristic model could outperform the normative and local computations models. If follows a logical rule: if during an intervention on variable X, variable Y crossed a value threshold, then a causal event is detected and the model collects information about whether the signs of X and Y match. If $\text{sign}(X) = \text{sign}(Y)$, the model considers it as evidence for a positive link, else it treats it as evidence for a negative link. To allow for the detection of weak and strong links, we add a temporal rule which states that if the threshold is crossed before a specific time threshold, then treat the link as strong, else treat it as weak. Note that this assumes that link strength has a direct effect on the speed of manifestation of the effect. The final judgement is given by the link values for which most

evidence was collected during a trial. By virtue of its simplicity the CEA model is blind to prior knowledge.

Parameters The CEA model is fit with a *causal event threshold* and a *guess* parameter, the former defining the value at which a causal event is detected and the latter the probability of answering anything else than the graph with the most collected evidence. We also fit a version with the added *time threshold* parameter to distinguish between weak and strong links.

3.4.4 Temporal and Temporal Link-focused Local Computations

Finally, we propose to fit two Bayesian models satisfying our assumptions that participants learn by using interventions as anchors for inference and by only updating outgoing links from the variable they are intervening on. For these, we forego the assumption of the CEA that inference is done using a threshold to detect discrete causal events and simply assume Bayesian learning to focus on how data is selected rather than processed. We propose two dimensions for locality: *temporal locality*, where interventions serve as beacons of focus, and *link focus*, where only beliefs about outgoing links from the last held node are considered. At one end of this locality space, we have agents who update their beliefs about all links at every single time point, such as the normative or standard local computations models. At the other end, we have agents that only consider data during interventions and update a single link at a time. Given that we only know when and on which node participants intervened, we cannot assert whether they were updating both outgoing links simultaneously or sequentially and will thus not model up to that level of granularity. Notice that *link focus* is nested within *temporal locality* in that there cannot be a focus on outgoing links from a held node if no node is held. We hypothesise that both dimensions are necessary to best capture participants' judgement and to evaluate this claim, we propose to fit two models. A first model we call Temporal LC, which is equivalent to the standard local computations model but restricts the data it considers to data generated during an intervention. During any intervention, it updates its beliefs about all links but the links directed towards the held node, as a standard local computations model would. The update rule for a given link, ignoring known parameters, can be expressed as follows:

$$p(\Gamma_{ij}|x^j_{t+dt}, \mathbf{x}_t, a_t) \propto p(x^j_{t+dt}, |x^i_t, x^j_t, \Gamma_{ij})^{\mathbb{1}[a_t \neq \emptyset, a_t \neq j]} p(\Gamma_{ij}) \tag{11}$$

where $\mathbb{1}\left[a_t \neq \emptyset, a_t \neq j\right]$ is an indicator function returning 0 if no intervention is being performed, the assumption of the temporal LC model, or if the intervention is on the target variable j, from the definition of a causal intervention.

The second model we fit is called the Temporal Focused Local Computations (TF LC). It is equivalent to the Temporal Local Computations model but further restricts its updates to outgoing links from the held variable. The update rule for a given link, ignoring known parameters, can be expressed as follows:

$$p(\Gamma_{ij}|x^j_{t+dt}, \mathbf{x}_t, a_t) \propto p(x^j_{t+dt}, |x^i_t, x^j_t, \Gamma_{ij})^{\mathbb{1}[a_t = i]} p(\Gamma_{ij}) \tag{12}$$

where the indicator function $\mathbf{1}\left[a_t = i\right]$ only returns one when the agent is performing an intervention on node i,

ensuring that the model only updates its beliefs about outgoing links from i, the held node. Notice how both the Temporal and Temporal Focused LC only differ from LC in the condition under which they choose to update their beliefs, the update equation itself is unchanged.

Parameters In both cases, similarly to previously presented Bayesian models, we fit a τ temperature parameter applied to the posterior. We also fit versions with and without prior influence (τ_p).

3.5 Model fitting results and discussion

Table 2 summarises the results from fitting inference models for Experiment 1, 2 and 3. We fitted each model to each participant by finding the parameter values which maximise the log likelihood of participant's final judgements about the true causal graph given the data generated in each each trial. The parameters fit for each models are given in the first five columns of the table. Model selection was done both with Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC). These results are displayed in the last four columns of Table 2. Overall, we see that the Temporal & Focused Local computations model provides a better fit in both versions, i.e. with and without a prior peaked on participants' elicited prior graphs. While both show that the TF LC with prior is on average a better fit than other models, the AIC appears to favour the version with prior, being a better fit for more participants than the version without.

Bayesian inference outperforms causal event abstraction. While our adapted version, i.e. augmented for strength detection, of the Causal Event Abstraction model (CEA) proposed by Davis et al. (2020) and Rehder et al. (2022) best fit some participants suggesting that it may still be a heuristic used by some, we show that a Bayesian inference agent, the Temporal Focused Local Computations (TF LC) model, performs better overall. Recall that while the CEA model implies the focus on interventional data, it commits to a threshold-based evidence accumulation process, whereby evidence in favour of a link is gathered based on whether the target cause variable crosses a threshold value. Beyond the fact that this model is framework specific, i.e. tied to the experimental design, our results show clearly that a more generic Bayesian account of inference, augmented by the constraint to solely consider data during interventions, offers an overall better fit to participants' judgements. We therefore propose that the good performance of the CEA model previously observed by Rehder et al. (2022) stemmed mostly from this constraint rather than from its theoretical claim of a threshold based evidence accumulation process. Furthermore, the CEA model makes a very specific prediction about the interventions that participants should perform. It predicts that in all circumstances, participants should hold a variable at the bounds in order to maximise the chances of its potential effects to remain above the threshold as it is the only way for it to gather evidence. In a dedicated analysis of participants' interventions in the presented experiments, Btesh et al. (2023b) have shown that much like in previous research (Rehder et al., 2022), participants often hold variables at the bounds for extended periods of time. However, a more in-depth analysis revealed that they also regularly swipe from one end of the scale to the other. In short, participants hold a variable at an end of the range, then swipe it to hold it at the other end for

some time. This behaviour cannot be accounted for by the CEA, meaning that while some participants may be using such a heuristic, we cannot consider it a complete model of human learning in this setting. Interestingly, a normative Bayesian actor seeking to maximise the amount of generated information will tend to only swipe (see Appendix 6.1). Clearly, participants' behaviour interpolates between these both extreme cases, highlighting a gap in our current understanding of their active learning process.

Both temporal and focus on outgoing links are required to capture participants' behaviour Table 2 shows unambiguously that adding the constraint that an agent only updates its beliefs during interventions is not sufficient to better capture participants' judgements than standard LC. Indeed, the simple Temporal LC is systematically outperformed by the Temporal Focused LC, which further constrains belief updating to outgoing links from the current intervention variables. This finding is crucial in that it specifies that locality in continuous time causal learning for humans entails a temporal and structural focus where interventions are the guiding tool for inference. The two best performing models, i.e. TF LC and CEA w. strength, both fundamentally rely on this assumption.

Priors matter but not for everyone. We see that a significant number of participants were best fit with a model with a non-flat prior distribution (~~.31~~ 31% in experiment 2 and ~~.26~~ 26% in experiment 3), i.e. one peaked on their prior judgement g_p, suggesting that the errors in final judgements observed behaviourally were not random errors but resulting from a bias towards prior judgements. It is however clear that more participants were best fit by a flat prior, suggesting that prior elicitation did not affect everyone. Overall, the split was roughly even for participants best fit by TF LC. We will unpack the differences between these two groups later. From a Bayesian standpoint and given the amount of datapoints available, it is expected that the prior needs to be heavily peaked in order to affect posterior judgements. We will later provide evidence for the interpretation that priors are used by some participants as expert knowledge, helping them to correctly identify indirect effects. Finally, the effect of priors on posterior judgements provides further evidence against the CEA model for those participants, as by definition, it cannot represent prior knowledge. ~~Indeed, by virtue of being heuristic, the CEA is~~

Model comparison using euclidean distance to participants' judgements. Figure 7 A. shows how close to participants judgements each model's predictions are (higher values are closer and 1 would be an exact match). In accordance with the other modelling results, we can see that Temporal Focused LC with prior is closer to participants judgements than all others. We nuance this result by noting that none of the models is very close to participants' judgements, TF LC with prior only reaching on average .85. While TF LC may capture key principles of human learning through task decomposition and data selection, it rests on implausible assumptions. Specifically, it still assumes that participants are able to estimate parameter values, here parameters θ and variance σ of the OU network, simply from the training phase of the experiment. Relaxing the assumption that key parameters are simply known to participants could open avenues for bridging the gap in fit we observe in these studies but would require

using approximate methods for Bayesian inference such as Markov Chain Monte Carlo (Bramley et al., 2017b) or variational inference (Sanborn, 2017; Btesh et al., 2023a). Furthermore, the Causal Event Abstraction (CEA) model, by proposing that participants decompose continuous data into discrete causal events, provide a solution to the unlikely assumption that participants perform Bayesian updates. While the TF LC provides an overall better fit, it reintroduces this assumption and future research using approximation methods for Bayesian updates should focus on addressing this shortcoming.

Table 2: Model fitting results for experiments 1, 2 and 3

Models	τ	τ_p	CE threshold	Strength threshold	Guess	Mean BIC individual	Number best fit	Mean AIC individual	Number best fit
Experiment 1									
Baseline						77.25	4	77.25	6
Normative	7.658					55.36	5	57.98	3
Normative w. prior	7.661	272.9				56.74	0	57.97	1
LC	1.609					59.58	3	62.20	1
LC w. prior	1.610	450.5				60.59	1	62.20	1
CEA			.625		.437	58.90	0	60.14	0
CEA w. strength			.744	48.398	.400	62.02	16	63.91	17
Temporal LC	1.655					60.94	3	63.56	2
Temporal LC w. prior	1.650	355.0				61.88	1	63.12	2
Temporal Focused (TF) LC	2.158					**51.81**	**22**	54.43	**20**
TF LC w. prior	2.107	186.8				52.55	5	**53.79**	6
Experiment 2									
Baseline						77.25	3	77.25	5
Normative	6.378					61.02	17	63.64	13
Normative w. prior	6.378	398.2				62.26	0	63.50	2
LC	1.542					62.31	10	64.93	6
LC w. prior	1.561	427.8				62.11	1	63.35	3
CEA			.622		.503	64.95	0	66.19	0
CEA w. strength			.634	47.269	.467	65.93	15	67.79	14
Temporal LC	1.457					64.41	6	67.03	2
Temporal LC w. prior	1.537	550.0				63.56	6	64.80	11
Temporal Focused (TF) LC	1.977					55.67	**32**	58.29	26
TF LC w. prior	1.991	192.0				**54.32**	31	**55.55**	**39**
Experiment 3									
Baseline						96.57	9	96.57	10
Normative	7.531					86.66	2	89.05	0
Normative w. prior	7.963	116.6				88.03	1	88.81	2
LC	1.699					74.00	7	76.39	5
LC w. prior	1.706	313.5				74.93	2	75.69	3
CEA			.643		.470	77.70	0	78.48	0
CEA w. strength			.646	48.193	.454	73.14	28	74.34	27
Temporal LC	1.540					78.44	2	80.84	1
Temporal LC w. prior	1.546	84.1				78.48	1	79.26	3
Temporal Focused (TF) LC	2.084					65.04	**39**	67.43	28
TF LC w. prior	2.060	188.2				**64.44**	27	**65.23**	**39**

Note: Mean estimated parameter values for each model and each experiment. Mean participant level BIC and AIC and associated number of participants best fit by each model for each criterion.

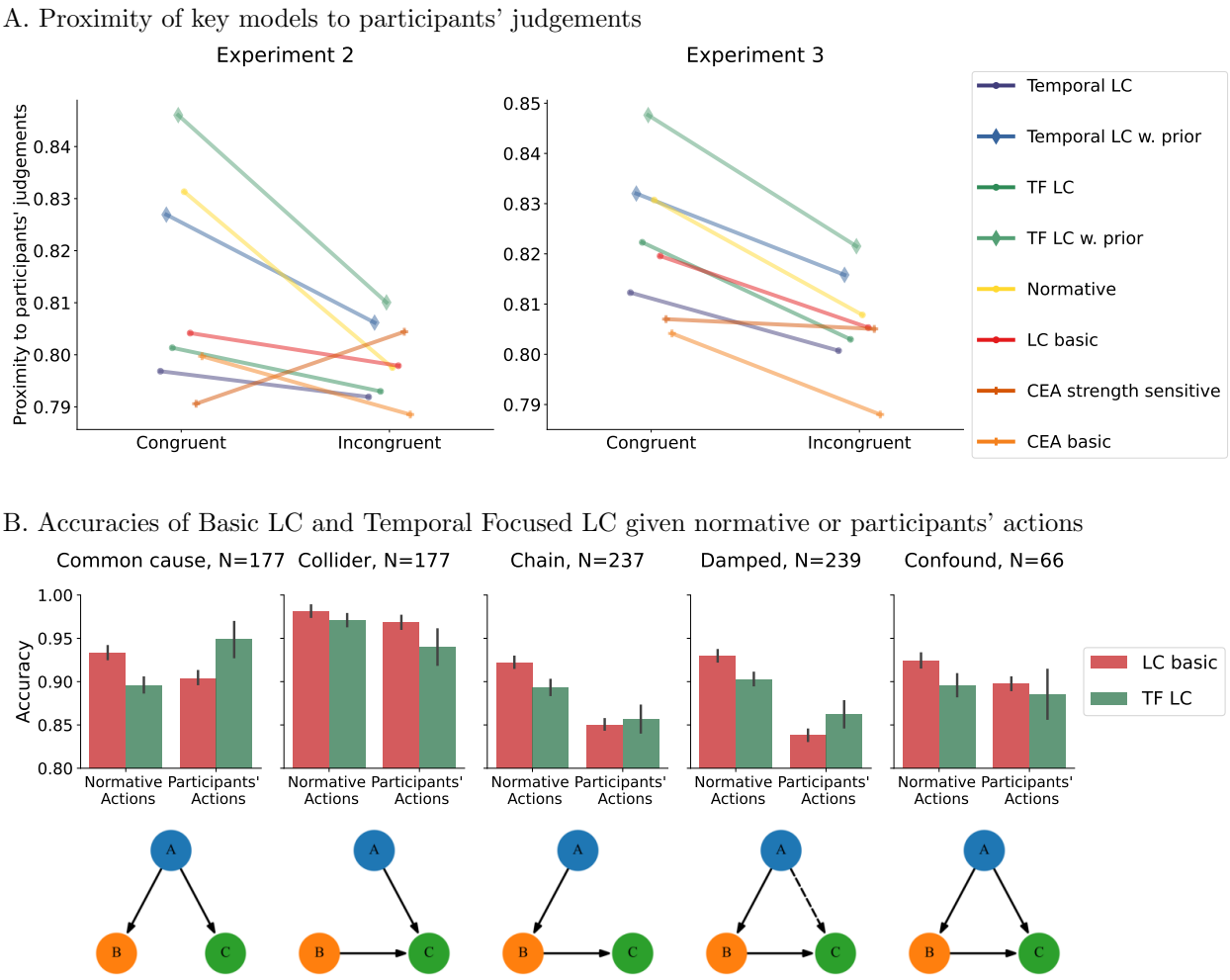


Figure 7: A. Proximity to participants' posterior judgements for each key computational model. B. Accuracies for basic LC and Temporal Focused (TF) LC when being trained using normative actions (see Appendix 6.1) or participants' true action for key benchmark graph structures. The latter are shown in the second row where full arrows are positive links and dashed arrows negative links.

On the efficiency of interventional data for inference. When comparing the ~~the~~ average performance of the Temporal Focused (TF) Local Computations model at recovering causal graphs (experiment 2: $.86 \pm .12$, experiment 3: $.88 \pm .11$) with the basic Local Computations model (experiment 2: $.87 \pm .07$, experiment 3: $.89 \pm .08$), we observe that they achieve a similar accuracy. There was no significant difference in mean accuracy in experiment 2 ($T = .71, p = .48$) nor in experiment 3 ($T = .58, p = .56$). This is striking. The average proportion of each trial spent intervening was .44 ($\pm .19$). The TF LC model only considers data during interventions meaning that at most, the TF LC achieved similar or better performance than the basic LC model, which considers all data, by only using half the available data. This illustrates how information efficient only focusing on interventions data is and suggests that participants' are using an adapative strategy. To investigate whether this remarkable performance depends on the choice of interventions made by participants, we used the optimal actor model (see Appendix 6.1) to generate

information optimal actions for the TF LC and LC basic models for all graphs across all four experiments. Figure 7 B. compares the accuracy of both models given optimal interventions or using participants' true interventions on benchmarks generic graphs. These results show that while quite efficient, the TF LC is always worse than the LC basic with optimal actions, which is expected given that in the optimal setting, it is never adaptive to not intervene, thus both agents use the same amount of data but the TF LC gets less updates as it only updates outgoing links from the currently held node. With human actions however, interventions are sparser and the Temporal Focused LC agent does generally as well if not better than the LC basic. We suspect that by only using interventional data, the TF LC agent only focuses on less ambiguous data points, as all incoming links into the held nodes are controlled for. At the same time, it saves computational cost by not attempting to update beliefs when not intervening. To conclude, the TF LC model not only best captures participants judgements but its inference process is best served, relative to the LC basic, by participants' actual interventions. This provides suggests that participants' inferences work in tandem with their action selection process to be as precise as possible while minimising the amount of data needed to achieve this performance.

4 Interpreting participants' performance in the light of modelling results

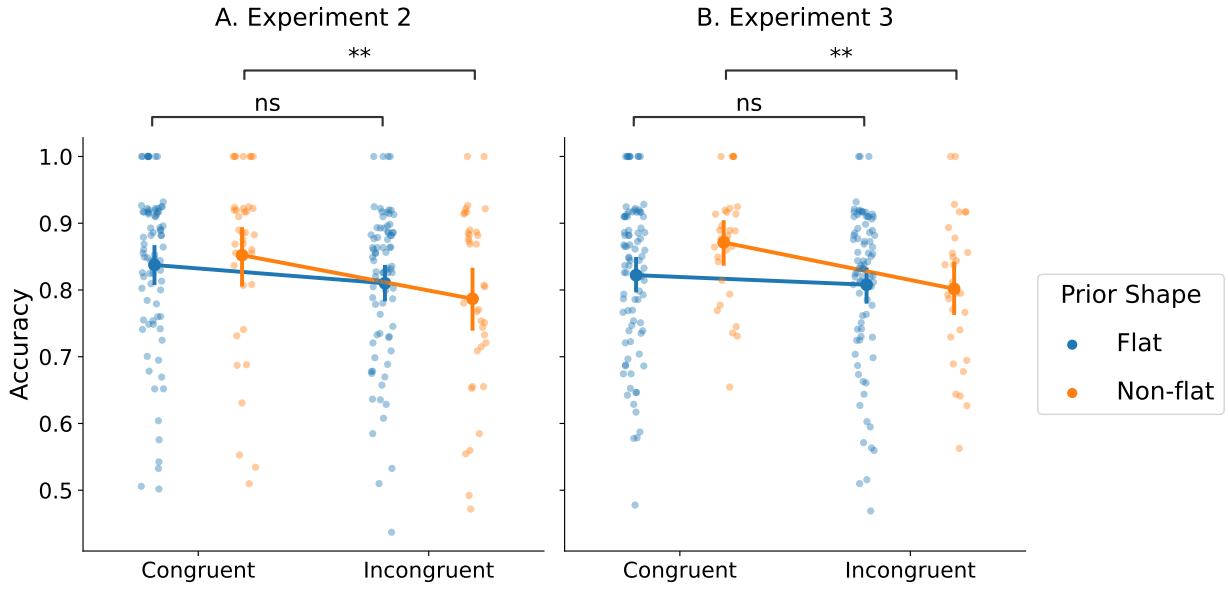


Figure 8: Mean differences in accuracy between condition for participants best fit by TF LC with a non-flat prior parameter and those best fit with assuming a flat prior and interactions with trial congruence for experiment 2 and 3.

Modelling participants' behaviour indicated that .31 of participants in experiment 2 and .26 of participants in experiment 3 were best fit by a flat prior rather than a prior peaked on their elicited prior graph. We show in

Figure 8 that when, using a variable we call *prior shape*, we split the sample into those who were best fit by a model with a non-flat prior and those that were not, we see interesting patterns in the dependencies between accuracy and prior judgements. For experiment 2 and 3, we ran linear mixed effect models predicting *accuracy* from *congruence*, *scenario*, *prior shape*, pairwise interactions and *local computation agent score*. In experiment 2, the fixed effect of *congruence* was significant ($F(114.81) = 8.52, p < 0.01$) and stronger than in the base analysis. There was a significant interaction between *prior shape*, *scenario* and *congruence* ($F(114.76) = 5.06, p < .05$) and a significant interaction between *prior shape* and *scenario* ($F(115.47) = 6.02, p < .05$) whereby participants best fit with a peaked prior did better at congruent trials when it was in the *crime* scenario than when it was in the *finance*. Overall, accuracy across congruent trials was higher in the group best fit by a prior peaked on their judgement ($.0724, T = 2.650, p < .01$) and no significant effect for the group best fit with a uniform prior. The fixed effect of *local computations agent score* was significant ($F(222.39) = 7.23, p < .01$). This measure is generated by having a local computations model perform inference on each graph while selecting optimal actions (see Appendix 6.1) and is used as a control for the impact of indirect links on the recoverability of the graph. This suggests that the graph generated in a scenario, here *finance*, tended to be harder to recover than the other explaining the interaction. Experiment 3 shows a clearer pattern, the fixed effect of *congruence* was also stronger with this model ($F(112.40) = 8.82, p < 0.01$) than with the one where all participants were pooled together. The only significant interaction was between *congruence* and *prior shape*, whereby only participants best fit with a prior peaked on their judgement showed a significant accuracy difference between congruent and incongruent trials ($.0743, T = 2.887, p < 0.01$). The full results of the fitted mixed effect regressions are in Supplementary Materials 5. This suggests that manipulating *congruence* strongly affected some participants, but others to a lesser extent. Moreover we can see that participants best fit by a model with a non-flat prior performed better in congruent trials relative to other participants, who did not show a performance difference between the conditions. In figure 7 A., we observe the same pattern in the Adaptive and TF LC models with prior, where they perform better in congruent trials than the one without and are both as accurate in incongruent trials.

Domain specific priors as expert knowledge used for detecting ambiguous mediators The chain bias has been a staple of causal structure learning experiments for more than two decades and describes the tendency of participants to incorrectly interpret an indirect effect as a direct one (Lagnado and Sloman, 2002, 2004; Steyvers et al., 2003; Fernbach and Sloman, 2009; Bramley et al., 2017a; Davis et al., 2020; Rehder et al., 2022). This mistake can lead to either reporting a nonexistent link or failing to report a direct link in cases where the indirect effect cancels the existing direct effect. Often, disambiguating between a direct and indirect effect is extremely difficult given the tools available to participants in experiments, e.g. only able to intervene on a single variable or only being able to observe rather than intervene. Usually participants prefer the option that is intuitively most consistent with their immediate observations, leading to systematic errors. In short, the chain bias stems from failing to detect a mediator variable in a causal relationship. However, in cases where they can lean on prior knowledge and where this knowledge is consistent with the ground truth, we propose that this helps them to correctly assess an ambiguous pattern in the data. To test for this, we create a measure which counts the number of errors in judgements made

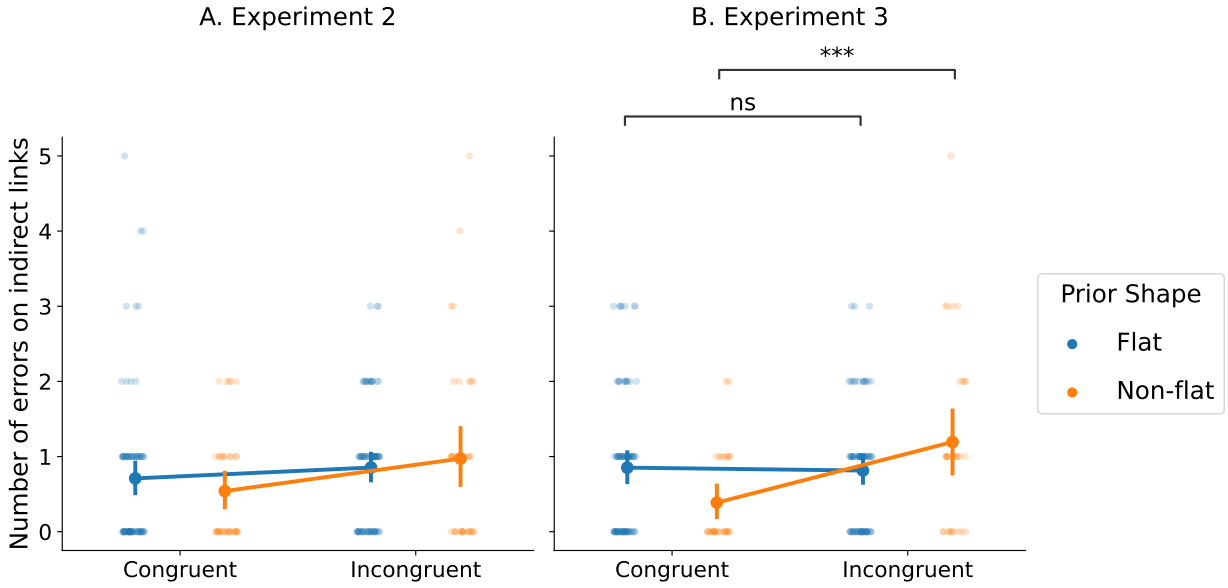


Figure 9: Mean number of errors made due to indirect links for each trial congruence and participants best fit with flat and non-flat priors.

on links where an indirect effect led to ambiguous data, essentially counting the number of errors due to a chain bias for each trial. We then fit linear mixed effect models for experiment 2 and 3 predicting this number of errors from whether the participant making them is best fit with a non-flat prior, i.e. *prior shape*, and the experimental condition, i.e. *congruence*. To make sure that any observed effect does not stem from participants in the non-flat prior group simply generating graphs with less indirect effects, we control for the number of indirect effects in the graphs generated by both groups for experiment 2 and 3. The full results are in Supplementary Materials 6 and summarised in figure 9. While unsurprisingly, the number of indirect links in the graphs strongly predicts the number of errors (experiment 2: $F(237) = 69.46, p < .001$, experiment 3: $F(231) = 71.49, p < .001$), we can see that for experiment 2, there was a unique fixed effect of *congruence* ($F(231) = 9.74, p < .01$) where participants in congruent trials made less errors on links affected by indirect effects. Despite a trend, interaction between *prior shape* and *congruence* was not significant. Again, the findings in experiment 3 were clearer. There was a significant fixed effect of *congruence* ($F(231) = 7.19, p < .01$) and a strong significant interaction between *congruence* and *prior shape* ($F(231) = 11.99, p < .01$). Participants best fit with a non-flat prior made on average .718 less errors in congruent trials than the rest of the sample ($t = -3.58, p < .01$). This was not the case in incongruent trials.

In short, we see that leaning on the prior leads participants to be less prone to the chain bias. This shows that while active learning in such dynamical settings is best represented with the local computations model, prior knowledge can still be integrated to allow participants to keep in mind a larger subset of links and in turn be more precise in their judgement than participants who discard their prior.

5 General Discussion

5.1 Priors consistent with the ground truth reduce biases stemming from learning with local computations

Complementing other recent accounts of causal structure learning in continuous time (Bramley et al., 2018, 2017b; Rehder et al., 2022; Gong et al., 2023), we show that while people are quite capable learners, they are still prone to systematic errors. We found evidence for hypothesis 1: that these errors would reflect a local computations account of causal learning, namely that participants would mistake indirect effects for direct ones. However, we showed that the interplay between prior knowledge and inference could help some participants circumvent these systematic errors.

In experiment 1, we replicated the classical chain bias, whereby participants erroneously report a direct link between the first and any indirect descendant in a causal chain, but also show that in the case where the indirect effect is preventative, i.e. a indirect effect counteracts a genuine direct effect, participants fail to report the direct link. We note that the bias does not solely lie in reporting or failing to report a direct link but rather in the preference participants have for the perceptually evident case. We however nuance this bias by demonstrating that when consistent with the ground truth, prior beliefs can serve as an anchor to correctly interpret such cases of ambiguous data, showing that participants are in fact able to recover the correct structure.

We showed that encouraging the use of domain-specific priors by providing informative labels for the variables in a causal system shaped participants' judgments. We observed that on average participants performed better when their prior was closer to the ground truth graph. This expected finding is in line with previous research on domain-specific priors in causal reasoning and provides evidence for hypothesis 3. We note that the observed effect was modest, and not consistent across all participants. This may be due to the design of our task not satisfying the requirements for theory-based causal induction (Griffiths and Tenenbaum, 2009). That is, even with informative labels, the task is still abstract compared to the rich and complex environments we encounter in daily life, so some participants may not have fully bought into the match between the task variables and their real world counterparts. Despite this, some participants still leaned on their prior in their learning, leading them to on average over perform in congruent trials and, to a lesser extent, under perform in incongruent trials relative to generically labelled trials. Across experiment 2 and 3 about a third of participants were affected by the prior manipulation. The rest, unencumbered by their priors, seemed to treat the task as a game rather than focusing on the meaning the labelling of the variables might carry. While the trends are subtle in this set of studies, we suspect that the abstract nature of our task under-represents how such mechanisms are harnessed in daily life. For instance, a computer scientist will be more likely to accurately diagnose the reason that a given application makes a computer crash from their knowledge of its interactions with the rest of the computer's structure (e.g. a known memory allocation bug), while a lay person may have to stick with a simpler, less useful but more perceptually evident explanation: opening this application causes the computer to crash. This pattern of results shows that the chain bias is not fatal. Participants are able, with relevant structural knowledge, to offset the errors stemming from their local focus.This interpretation of the role of priors falls well within

the literature on domain specific priors, i.e. expert-like knowledge, which can be leveraged to correctly find mediators from ambiguous evidence. There may, however, be a more computational account of the observed weight of priors working in tandem. As the data are abundant and heavily correlated and therefore definitely not independent and identically distributed, participants may be devaluing datapoints in their posterior estimate in order to account for this and save computational resources. The combination of this effect and the effect of priors as knowledge can be descriptively modelled by peaked priors, as we did here, but future research should aim to disambiguate both as they represent theoretically distinct phenomena.

5.2 Locally focused learning as an adaptive task decomposition strategy

Consistent with our prediction in hypothesis 2., in continuous time and continuous state space, where data are abundant, participants are best fit by a computational model that adaptively selects sections of available data for inference. The main guide for this selection appears to be actions, whereby participants mostly consider data during their interventions but also data which provides information about the direct outgoing links from the intervened node. The data shows clearly that both those components are needed to best fit participants' judgements. Simply put, participants seem to look *when* they do and close to *what* they do and ignore the rest of the data (Lagnado and Sloman, 2004). This focus on interventions shows that when given the opportunity to experiment with a system themselves by using interventions, participants not only take it but structure their whole learning process around it, reinforcing the view that humans are fundamentally active learners. Furthermore, this allows them to decompose the task into sub goals which are more tractable and offer a clear sequence of actions to take: intervene on one variable, learn about its effects and repeat until the full structure is covered. Such decompositions have been well documented by Bramley et al. (2018) and more recently by Correa et al. (2022) and allow for less complex updates at the cost of accuracy. Finally, this latter point paired with the frugal intake of information only focusing at interventional data implies, i.e. rarely considering more than half of the available data, is reminiscent of theoretical framework of bounded rationality (Simon, 1955) and its later development, resource-rational cognition (Lieder and Griffiths, 2020), where inference and planning are represented as trade-off processes between maximising utility and minimising computational cost. To be precise, by showing that in addition to temporarily simplifying the causal structure, interventions allow agents to push the system ~~in~~ into extreme and unlikely states providing less ambiguous and thus more informative data, this work further consolidates a view of interventions as a fundamental tool for frugal but efficient learning strategies. We believe that a proper resource rational analysis is warranted in future research to properly investigate this question.

5.3 On the necessity of representing discrete causal events

As stated above, the assumptions of the Temporal Focused Local Computations (TF LC) Bayesian model are two-fold: only consider data generated during interventions and only update beliefs about the links outgoing from the intervened variable. While not constituting the main postulate behind the Causal Event Abstraction model proposed by Rehder et al. (2022), these are implied by it as they are necessary to its functioning. Indeed, contrary to a Bayesian

approach, the CEA proposes that participants discretise events using a heuristic whereby evidence is accumulated about the presence and the sign of a link when, given an intervention on a variable A, a second variable B crosses a threshold value. This triggers the recording of a causal event. All events are then combined at the end of the trial to provide the final judgement. The crucial difference between CEA and TF LC is therefore about the credit given to the threshold-based inference process assumed by the former.

Our stance, given TF LC being the best fitting model across all experiments, is that a Bayesian model considering the true data stream without further processing and equipped with those two assumptions about locality is more potent at capturing the behaviour of most participants. First, the fact that some participants relied on prior knowledge to avoid mistakes on indirect links suggests that participants conditioned their learning on previous information. Such a behaviour cannot easily be accounted for by the heuristic policy which underpins the Causal Even Abstraction model. Furthermore, the CEA predicts that participants should learn only by holding variables at the bounds to maximise chance of triggering a causal event. To the contrary, an ideal Bayesian intervener, i.e. one attempting to maximise the information generated, would constantly swipe up and down. Participants fall between those two extreme cases, i.e. only holding or only swiping: they do hold but also swipe from one end to the other, only to hold again (Btesh et al., 2023b). Swiping is a behaviour cannot be accounted for by CEA's inference process but is perfectly predicted by a simple Bayesian model considering unprocessed data, thus further favouring TF LC.

However, CEA was a better fit for a stable subset of participants across experiments and thus cannot be discarded so easily. While both CEA and TF LC assume unrealistic computations, the ones assumed by TF LC are particularly so in that they assume a single Bayesian update per frame while the CEA simply counts the frame for which the threshold is crossed. Beyond Bayesian updates, TF LC stays silent about how the continuous signal is processed into information used for inference. Allowing the model to consider unprocessed data still leaves a gap in our understanding of participants' behaviour. If people, much like our Bayesian model, truly had the ability to integrate raw data, they would never need to hold variables for extended periods of time, which has been shown to be a core feature of participants' policies ~~Rehder et al., 2022; Btesh et al., 2023b~~(Rehder et al., 2022; Btesh et al., 2023b). We fully recognise this shortfall and bridging this gap will involve seriously considering how the continuous data is processed by participants. We consider the TF LC to be a first step in this direction. The better fit of TF LC and the fact that interventions performed by participants are still more consistent with a Bayesian learner should encourage us to pursue approximate Bayesian methods to bridge this gap rather than task specific heuristics. These methods have proven to be successful in modelling participants' behaviour (Dasgupta et al., 2020; Zhu et al., 2020; Sanborn, 2017). To be clear, our reservation does not lie in discretisation per se as it has already been shown to be a plausible causal learning strategy ~~Gong et al., 2023~~(Gong et al., 2023). Our doubts lie in the specific heuristic of a causal event detection threshold and how these events are then integrated. These are contrary to most participants' interventions policies do not allow the model to represent prior beliefs. Finally, while it is probable that different people use different strategies, one of which may be threshold based while another using a different processing method, in both cases, the baseline premises required to capture participants' judgement remain constant: consider mostly

interventional data and focus on direct link outgoing from the variable on which they intervene.

5.4 Limitations

A key theoretical limitation of this work lies in the fact that the best performing models, i.e. Temporal Focused Local Computations and Causal Event Abstraction, are incapable of learning from observational data. Extending this claim to human participants is simply not reasonable as there is ample evidence that observational learning learning does occur, at least in suitably simple circumstances (Valentin et al., 2023; Rehder, 2017; Lagnado et al., 2007). While our results make clear that participants have a strong preference for interventional data when given the opportunity to use it, we do not claim that humans are as strict as the TF LC or the CEA models. Future research should attempt to model a softer evidence weighting process. The model should be able to attend to observational data and represent its weight in its posterior beliefs relative to interventional data. Furthermore, as we have discussed above, the question of how data is processed remains to be addressed. Our best performing model, the TF LC, can observe and use unprocessed data, i.e. consider each new frame one at a time. This possibility is psychologically unrealistic and it is likely that some simplification strategy is happening in order to manage this abundance of data. Future research should focus on integrating our findings here, that participants rely primarily on interventional data, use a local computations approach and are able to mobilise prior beliefs, with a data processing approach respecting the computational constraints of the human mind. Such an approach may be rule based, for instance considering that a link going from A to B is positive if B increases when A does.

An assumption common to our normative analysis of the experiments, is that participants, rather than having to learn about them, hold some scalar estimates of the parameters (here θ and σ) which are not causal link values in the generative model. Davis et al. (2020) have proposed that these are noisy estimates learned from the introduction to the task but they were not formally learned. This assumption comes from the fact that inference in a OU network where agents hold beliefs about all parameters is not tractable without making simplifying assumptions, e.g. mean-field approximation, about the factorisation of the joint distribution. Future research should aim at comparing the performance in matching participants judgements of approximations methods such as MCMC or variational Bayes to existing models. MCMC-like methods have, in discrete time settings, yeilded insights into human learning patterns (Bramley et al., 2017a; Davis et al., 2020). Extending these findings to continuous time is a natural continuation of this research project.

These studies exemplify that exploring general effects of individual priors can be challenging. One reason is that making the task materials dependent on each participant required us to pick a rule for manipulating the graphs to produce incongruent ground truths. The generative process used in experiment 2 and 3, while balanced in terms of the number of connections, still allows for the generation of graphs which are not Markov equivalent to the prior, which may affect learning outcomes in ways that we may not be fully able to control for. It is therefore a complex balance to find between making the ground truth significantly different from the prior judgement while controlling for systematic effects on difficulty.

Finally, in daily adult life, it is much more common for individuals to learn about causal systems with the end goal of controlling them: learning how to ride a bicycle, drive a car, use a new smartphone, play a video game. The Ornstein-Uhlenbeck network framework, as developed by Davis et al. (2018), lends itself well to simulating and modelling dynamic control problems. While the present studies provide an in-depth exploration of the setting where learning the structure of the environment is the end goal, the question remains whether the behaviours observed here would extend to control settings. As such an experimental setting where the goal is to maintain the system in a given state rather than explicitly learn its dynamics would be a meaningful path to understand the interplay between human control and learning.

5.5 Conclusion

In three experiments, we found evidence for the influence of prior knowledge on causal attributions in an active dynamical structure learning task. We showed that this effect is not consistent across participants and that those who did rely on prior beliefs in their inference performed better when the ground truth was consistent with their beliefs, but worse when it was not. We showed that the boost in performance in trials where the ground truth was consistent with participants' prior beliefs was partly due to a protection from the chain bias: participants were less likely to misrepresent indirect connections as direct.

We showed that an adapted version of the local computations model provided the best account of participants' behaviour. This new version, Temporal Focused Local Computations, adaptively considers data during intervention time and focuses on outgoing links from the intervened node. This supports the idea that participants, rather than use all available data, split the inference process into smaller, more tractable sub-problems, and use interventions as beacons for data selection. Further, we showed that such a model achieves equivalent performance to a version which considers all available information, while only using half the data. This adaptive strategy saves half of the computational cost associated with belief updating.

Humans are active learners, dexterously weaving observations, interventions and prior knowledge to solve problems or make inferences about the world. We have seen that not only are the strategies devised by humans resource efficient, they are also as accurate, assuming an agent which intervenes on the graph like a human, as much more data heavy strategies such as the standard local computations model. This importantly suggests that participants' inference process is best served by the types actions that they chose to perform, a notion which has been hinted at by Rehder et al. (2022). Future modelling efforts should first provide a more in-depth analysis of participants' action policies and then propose dual models of observations and actions which satisfy the constraints set by these policies. Our findings suggest that in this setting, humans are able to find cost effective strategies with relatively small performance trade-off. Future research should thus integrate the frugality implied by the temporal focused strategy employed here by participants with a resource rational framework to attempt to formalise the emergence of such behaviours.

Conflict of interest

The authors declare no conflict of interest regarding the present work.

References

Ahn, W.-k. and Dennis, M. J. (2000). Induction of causal chains. *Proceedings of the Annual Meeting of the Cognitive Science Society*, (22).

Bramley, N. R., Dayan, P., Griffiths, T. L., and Lagnado, D. A. (2017a). Formalizing Neurath's ship: Approximate algorithms for online causal learning. *Psychological Review*, 124(3):301–338.

Bramley, N. R., Gerstenberg, T., Mayrhofer, R., and Lagnado, D. A. (2019). Intervening in Time. In Kleinberg, S., editor, *Time and Causality across the Sciences*, pages 86–115. Cambridge University Press.

Bramley, N. R., Gerstenberg, T., Tenenbaum, J. B., and Gureckis, T. M. (2018). Intuitive experimentation in the physical world. *Cognitive Psychology*, 105:9–38.

Bramley, N. R., Lagnado, D. A., and 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., and Lagnado, D. A. (2017b). Causal learning from interventions and dynamics in continuous time. In *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*, volume 1, pages 150–155.

Btesh, V., Bramley, N. R., Fränken, J.-P., Speekenbrink, M., and Lagnado, D. A. (2023a). Variational inference in dynamical active causal learning. In *Proceedings of the Computational Cognitive Neuroscience Society Meeting 2023*.

Btesh, V., Bramley, N. R., Speekenbrink, M., and Lagnado, D. A. (2023b). Swipe and hold: Composing interventions in continuous time causal learning. In *Proceedings of the 45th Annual Meeting of the Cognitive Science Society*, volume 45 of *Proceedings of the Cognitive Science Society*. The Cognitive Science Society.

Coenen, A., Nelson, J. D., and Gureckis, T. M. (2019). Asking the right questions about the psychology of human inquiry: Nine open challenges. *Psychonomic Bulletin & Review*, 26(5):1548–1587.

Correa, C. G., Ho, M. K., Callaway, F., Daw, N. D., and Griffiths, T. L. (2022). Humans decompose tasks by trading off utility and computational cost. arXiv:2211.03890 [cs].

Dasgupta, I., Schulz, E., Tenenbaum, J. B., and Gershman, S. J. (2020). A theory of learning to infer. *Psychological Review*, 127(3):412–441.

Davis, Z. J., Bramley, N. R., and Rehder, B. (2020). Causal Structure Learning in Continuous Systems. *Frontiers in Psychology*, 11:244.

Davis, Z. J., Bramley, N. R., Rehder, R. E., and Gureckis, T. M. (2018). A Causal Model Approach to Dynamic Control. *Proceedings of the 41st Annual Meeting of the Cognitive Science Society*, pages 281–286.

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

Gershman, S. J. (2015). A Unifying Probabilistic View of Associative Learning. *PLOS Computational Biology*, 11(11):e1004567.

Gershman, S. J. (2017). Context-dependent learning and causal structure. *Psychonomic Bulletin & Review*, 24(2):557–565.

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

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., and Danks, D. (2004). A Theory of Causal Learning in Children: Causal Maps and Bayes Nets. *Psychological Review*, 111(1):3–32.

Grice, G. R. (1948). The relation of secondary reinforcement to delayed reward in visual discrimination learning. *Journal of Experimental Psychology*, 38(1):1–16. Place: US Publisher: American Psychological Association.

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

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

Guez, A., Silver, D., and Dayan, P. (2013). Scalable and Efficient Bayes-Adaptive Reinforcement Learning Based on Monte-Carlo Tree Search. *Journal of Artificial Intelligence Research*, 48:841–883.

Hagmayer, Y., Meder, B., Osman, M., Mangold, S., and Lagnado, D. A. (2010). Spontaneous Causal Learning While Controlling A Dynamic System. *The Open Psychology Journal*, 3(2):145–162.

Holyoak, K. J. and Cheng, P. W. (2011). Causal Learning and Inference as a Rational Process: The New Synthesis. *Annual Review of Psychology*, 62(1):135–163.

Jones, M. and Love, B. C. (2011). Bayesian Fundamentalism or Enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*, 34(4):169–188.

Kemp, C., Goodman, N. D., and Tenenbaum, J. B. (2010). Learning to Learn Causal Models. *Cognitive Science*, 34(7):1185–1243.

Kramers, H. A. (1940). Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7(4):284–304.

Kruschke, J. K. (2006). Locally Bayesian learning with applications to retrospective revaluation and highlighting. *Psychological Review*, 113(4):677–699.

Lagnado, D. A. and Sloman, S. (2002). Learning Causal Structure. In Gray, W. D. and Schunn, C. D., editors, *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society*, pages 560–565. Routledge.

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

Lagnado, D. A. and 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., Waldmann, M. R., Hagmaye, Y., and Sloman, S. A. (2007). Beyond Covariation: Cues to Causal Structure. In Gopnik, A. and Schulz, L., editors, *Causal Learning*, pages 154–172. Oxford University PressNew York, 1 edition.

Lieder, F. and Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, 43:e1.

Lohnas, L. and Kahana, M. (2014). Compound cuing in free recall. *Journal of experimental psychology. Learning, memory, and cognition*, 40(1). Publisher: J Exp Psychol Learn Mem Cogn.

Mandelbrot, B. and Taylor, H. M. (1967). On the Distribution of Stock Price Differences. *Operations Research*, 15(6):1057–1062. Publisher: INFORMS.

Newcombe, N. S., Sluzenski, J., and Huttenlocher, J. (2005). Preexisting Knowledge Versus On-Line Learning: What Do Young Infants Really Know About Spatial Location? *Psychological Science*, 16(3):222–227.

Pearl, J. (1997). *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, 1st editio edition.

Pearl, J. (2009). Causal inference in statistics: An overview. *Statistics Surveys*, 3(September):96–146.

Phillips, L. D. and Edwards, W. (1966). Conservatism in a simple probability inference task. *Journal of Experimental Psychology*, 72(3):346–354. Place: US Publisher: American Psychological Association.

Pilditch, T. D., Fenton, N., and Lagnado, D. A. (2019). The Zero-Sum Fallacy in Evidence Evaluation. *Psychological Science*, 30(2):250–260.

Rehder, B. (2014). Independence and dependence in human causal reasoning. *Cognitive Psychology*, 72:54–107.

Rehder, B. (2017). Reasoning With Causal Cycles. *Cognitive Science*, 41:944–1002.

Rehder, B., Davis, Z. J., and Bramley, N. (2022). The Paradox of Time in Dynamic Causal Systems. *Entropy*, 24(7):863.

Rehder, B. and Waldmann, M. R. (2017). Failures of explaining away and screening off in described versus experienced causal learning scenarios. *Memory & Cognition*, 45(2):245–260.

Rothe, A., Deverett, B., Mayrhofer, R., and Kemp, C. (2018). Successful structure learning from observational data. *Cognition*, 179:266–297.

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

Sanborn, A. N. (2017). Types of approximation for probabilistic cognition: Sampling and variational. *Brain and Cognition*, 112:98–101.

Schooler, L. J. and Anderson, J. R. (1997). The Role of Process in the Rational Analysis of Memory. *Cognitive Psychology*, 32(3):219–250.

Schulz, L. E. and Gopnik, A. (2004). Causal learning across domains. *Developmental Psychology*, 40(2):162–176.

Settles, B. (2012). *Active Learning*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Springer International Publishing, Cham.

Simon, H. A. (1955). A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, 69(1):99.

Sloman, S. A. and Lagnado, D. (2015). Causality in Thought. *Annual Review of Psychology*, 66(1):223–247.

Sobel, D. M., Tenenbaum, J. B., and Gopnik, A. (2004). Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cognitive Science*, 28(3):303–333.

Sobel, D. M., Yoachim, C. M., Gopnik, A., Meltzoff, A. N., and Blumenthal, E. J. (2007). The Blicket Within: Preschoolers' Inferences About Insides and Causes. *Journal of cognition and development : official journal of the Cognitive Development Society*, 8(2):159–182.

Speekenbrink, M. and Shanks, D. R. (2010). Learning in a changing environment. *Journal of Experimental Psychology: General*, 139(2):266–298.

Stengård, E., Juslin, P., Hahn, U., and Van Den Berg, R. (2022). On the generality and cognitive basis of base-rate neglect. *Cognition*, 226:105160.

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

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

Tversky, A. and Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157):1124–1131. _eprint: https://www.science.org/doi/pdf/10.1126/science.185.4157.1124.

Uhlenbeck, G. E. and Ornstein, L. S. (1930). On the Theory of the Brownian Motion. *Physical Review*, 36(5):823–841.

Valentin, S., Bramley, N. R., and Lucas, C. G. (2023). Discovering Common Hidden Causes in Sequences of Events. *Computational Brain & Behavior*, 6(3):377–399.

Waldmann, M., editor (2017). *The Oxford Handbook of Causal Reasoning*. Oxford Library of Psychology. Oxford University Press, Oxford, New York.

Waldmann, M. R., Cheng, P. W., Hagmayer, Y., and Blaisdell, A. P. (2008). Causal learning in rats and humans: A minimal rational model. In *The Probabilistic Mind:*, pages 449–479. Oxford University Press, Oxford.

Waldmann, M. R. and Hagmayer, Y. (2001). Estimating causal strength: the role of structural knowledge and processing effort. *Cognition*, 82(1):27–58.

Zhu, J.-Q., Sanborn, A. N., and Chater, N. (2020). The Bayesian sampler: Generic Bayesian inference causes incoherence in human probability judgments. *Psychological Review*, 127(5):719–748.

6 Appendix

6.1 An account of normative information driven action selection using tree search

We propose an account of normative action selection which is based on a tree search over sequences of actions where the value of a sequence, or policy, is given by the approximate expected information gained of that policy under current posterior beliefs. This method takes inspiration from Guez et al. (2013), where the current posterior is used to sample possible ground truth states of the world and use those to simulate future data and use these to evaluate action values. Figure 10 shows examples policies from such an agent with a normative Bayesian inference agent. The normative Bayesian agent recovers the ground truth graph in less than ten datapoints. We can see that the Bayesian agent iteratively pegs each variable to the extreme of the bounds. It follows that from a normative Bayesian standpoint, the maximisation of information gained is a constant swiping process at the bounds of the range.

Measuring action values Here as the task is solely information driven, action values $V_t(a)$ are simply given by expected information gained, which has been widely used in previous research (Bramley et al., 2015, 2018; Gong et al., 2023). The value of an action a at time t is given by

$$V_t(a) = \mathbf{H}(\mathcal{G}|a)_{gain} = \mathbf{H}(\mathcal{G})_{t_1} - \mathbf{H}(\mathcal{G}|a)_{t_2} \tag{13}$$

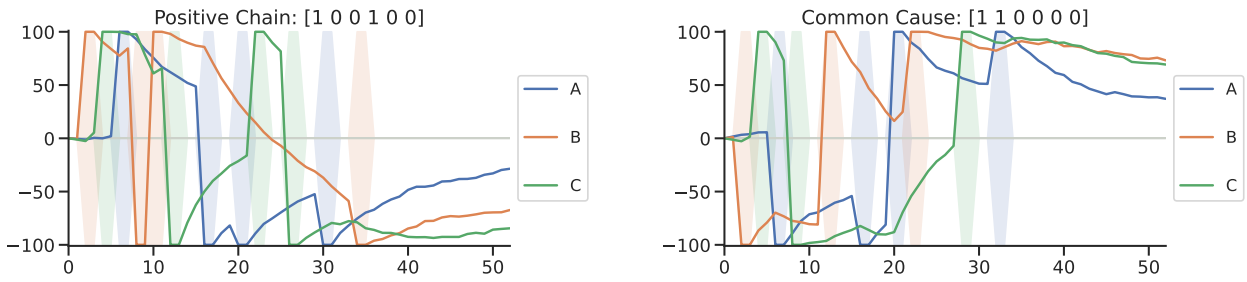


Figure 10: Example of actions taken by an normative Bayesian agent using normative information driven action selection. Each action lasts for at least one frame and are represented by the shaded areas. Actions stop when the agent reaches an arbitrary certainty threshold, here when the entropy of the posterior distribution reaches 0.01.

where $\mathbf{H}(\mathcal{G})_t$ is the entropy of the posterior over graphs at time t. It is thus simply the difference in entropy between t_1 and t_2 given that the agent took action a.

Managing a continuous state space A normative agent should have perfect information and be able to explore the whole space of interventions to find the optimal choice of sequence. However, it is unrealistic as the space of all interventions is continuous and even when discretised, it is still an intractable space to explore. This is especially true since the value of an intervention in a continuous but limited time depends not only on the expected gain in value, i.e. probability, utility or information, but also on the opportunity cost generated by the limited time of trial. Indeed, each datapoint spent intervening on one variable must not only maximise the gain but also minimise the loss of value of not intervening on another variable. This is a subtelty that we note but choose to ignore as this is not the focus of this paper. To reduce the space of possible actions, we assume that the agent discretise the space into bins. For instance, in the continuous interval $[-100, 100]$, the agent only considers the following values for possible actions: $\{-100, -75, -50, -25, 0, 25, 50, 75, 100\}$. We call r the size of this set where here $r = 9$.

Tree search algorithm To pick the policy which is estimated to be most informative, the agent samples a graph from its posterior at time t and simulates taking all possible actions at time $t+1$. These are the first leaves with each leaf is associated a gain in information. Then, for time $t+2$, the agent simulates taking all possible actions for each leaf in the tree, adding a new gain of information. The agents then repeats this process for a set number of graph sampled from its posterior and average the information gains over those. When the simulation ends, the agent can select the action which lies on the branch which is more informative on average.

We can use a recursive routine that simulates data for each graph sampled from $p(\mathcal{G}|\mathbf{x}_{1:t})$ by performing, for each leaf at depth d, all 1 step potential interventions, i.e. all $a \in A_d$, there are Kr such interventions, as each variable can be set to each of its possible values, observing new observation o_d and computing the information gained from the resulting posterior distribution. The routine stops when it reaches a specified depth parameter d, which controls the number of future simulated time steps. This routine is followed for C sampled graphs over which the information

gains are averaged. In all simulation, we chose $d = 2$ and $C = 5$.

Action selection from information gains Action selection from action values can be done in different ways. First, the agent can select an intervention deterministically, i.e. $\arg\min_{a \in A_t} R_C(a|g_s)$. This is an issue as the observations are noisy, i.e. normally distributed with variance dt. Also, interventions may be very similar in information gain for an agent performing normative Bayesian updating, therefore simply picking the intervention that minimise the estimated entropy reduction may not be the best course of action. A better and simple solution would be to randomly select an intervention $a \in A_d$ according to a softmax rule over the expected information gain $\mathbf{H}(G|a_{1:d}, o_{1:d})_{gain}$. For simplicity we denote the gain at depth d for interventions a_1 to a_d given a sampled graph g_s as $G(a_{1:d}|g_s)$, the gain between depth $d-1$ and d is thus denoted $G(a_d|g_s)$. The expectation over sampled graphs is $G_C(a_{1:d}) = \frac{1}{C}\sum_{c=1}^{C} G(a_d|g_s^c)$. The softmax rule is thus as follows:

$$p(a_{1:d}) = \frac{\exp(\theta G_C(a_{1:d}))}{\sum_a \exp(\theta G_C(a_{1:d}))} \tag{14}$$

where θ is an inverse temperature parameter, i.e. higher θ will favor more strongly options with higher negative expected entropies. Of course a ϵ-greedy policy can also be used to choose the following action.