Formalizing Neurath's Ship: Approximate Algorithms for Online Causal Learning

Neil R. Bramley
Peter Dayan
Thomas L. Griffiths
David A. Lagnado

Author note:

Neil R. Bramley, Department of Experimental Psychology, University College London; Peter Dayan, Gatsby Computational Neuroscience Unit, University College London; Thomas L. Griffiths, Department of Psychology, University of California Berkeley; David A. Lagnado, Department of Experimental Psychology, University College London.

This research was supported in part by an Economic and Social Research Council UK grant (RES 062330004).

Correspondence concerning this article should be addressed to: 201, 26 Bedford Way, University College London, London, UK,WC1H 0DS, Email: neil.bramley@ucl.ac.uk .

Experiment 2 previously appeared in a conference paper presented at The 37th Annual Conference of The Cognitive Science Society (Bramley, Dayan, & Lagnado, 2015). A preprint of the current manuscript was made available on arXiv on September 17th 2016 here: `https://arxiv.org/abs/1609.04212`.

Abstract

Higher-level cognition depends on the ability to learn models of the world. We can characterize this at the computational level as a structure-learning problem with the goal of best identifying the prevailing causal relationships among a set of relata. However, the computational cost of performing exact Bayesian inference over causal models grows rapidly as the number of relata increases. This implies that the cognitive processes underlying causal learning must be substantially approximate. A powerful class of approximations that focuses on the sequential absorption of successive inputs is captured by the *Neurath's ship* metaphor in philosophy of science, where theory change is cast as a stochastic and gradual process shaped as much by people's limited willingness to abandon their current theory when considering alternatives as by the ground truth they hope to approach. Inspired by this metaphor and by algorithms for approximating Bayesian inference in machine learning, we propose an algorithmic-level model of causal structure learning under which learners represent only a single global hypothesis that they update locally as they gather evidence. We propose a related scheme for understanding how, under these limitations, learners choose informative interventions that manipulate the causal system to help elucidate its workings. We find support for our approach in the analysis of three experiments.

KEYWORDS: active learning; causal learning; theory change; resource rationality; intervention

By adulthood, a normal person will have developed a sophisticated and structured understanding of the world. The "blooming buzzing confusion"(James, 2004, p462) of moment-to-moment sensory experience will have given way to a more coherent dance of objects and forces, relata and causal relationships. Such representations enable humans to exploit their physical and social environments in flexible and inherently model-based ways (Dolan & Dayan, 2013; Griffiths & Tenenbaum, 2007; Sloman, 2005). An important question, therefore is how people learn appropriate causal relationships from the data they gather by observing and manipulating the world. Much recent work on causal learning has used Pearl's (2000) *causal Bayesian network* framework to demonstrate that people make broadly normative causal inferences based on cues like observed contingencies or the outcomes of interventions, which are manipulations or tests of the system (e.g. Griffiths & Tenenbaum, 2009; Holyoak & Cheng, 2011; Lagnado & Sloman, 2004, 2006; Lagnado, Waldmann, Hagmayer, & Sloman, 2007). Related work has begun to explore how people engage in "active learning" – selecting interventions on variables in systems of interest in order to be effective at reducing uncertainty about the true causal model (Bramley, Lagnado, & Speekenbrink, 2015; Coenen, Rehder, & Gureckis, 2015; Sobel & Kushnir, 2006; Steyvers, Tenenbaum, Wagenmakers, & Blum, 2003).

Models of human causal learning based on Bayesian networks have tended to focus on what Marr (1982) called the *computational* level. This means that they consider the abstract computational problem being solved and its ideal solution rather than the actual cognitive processes involved in reaching that solution – Marr's *algorithmic* level. In practice the demands of computing and storing the quantities required for exactly solving the problem of causal learning are intractable for any non-trivial world and plausibly-bounded learner. Even a small number of potential relata permit massive numbers of patterns of causal relationships. Moreover, real learning contexts involve noisy (unreliable) relationships and the threat of exogenous interference, further compounding the complexity of normative inference. Navigating this space of possibilities optimally would require maintaining probability distributions across many models and updating all these probabilities whenever integrating new evidence. This evidence might in turn be gathered piecemeal over a lifetime of experience. Doing so efficiently would require choosing maximally informative interventions, a task which poses even greater computational challenges: consideration and weighting of all possible outcomes, under all possible models for all possible interventions (Murphy, 2001; Nyberg & Korb, 2006).

In order to understand better the cognitive processes involved in learning causal relationships, we present a detailed exploration of how people, with their limited processing resources, represent and reason about causal structure. We begin by surveying existing proposals in the literature. We then draw on the literature on algorithms for approximating probabilistic inference in computer science using these to construct a new model. We show that our new model captures the behavioral patterns using a scalable and cognitively plausible algorithm and explains why aggregate behavior appears noisily normative in the face of individual heterogeneity.

Many existing experiments on human causal learning involve small numbers of possible structures, semi-deterministic relationships and limited choices or opportunities to intervene.

These constraints limit the computational demands on learners, and thus the need for heuristics or approximations. Further, in most existing studies, subjects make causal judgments only at the end of a period of learning, limiting what we can learn about how their beliefs evolved as they observed more evidence, and how this relates to intervention choice dynamics. One exception is Bramley, Lagnado, and Speekenbrink (2015), which explored online causal learning in scenarios where participants' judgments about an underlying causal structure were repeatedly elicited over a sequence of interventional tests. Another is Bramley, Dayan, and Lagnado (2015), which built on this paradigm. Both papers explained participants' judgments with accounts that are not completely satisfying algorithmically, lacking cognitively plausible or scalable procedures that could capture the ways in which judgments and intervention choices deviated from the rational norms. Here, we develop the algorithmic level account and demonstrate that it outperforms or equals competitors in modelling the data from both previous papers and a new experiment.

The resulting class of algorithms embodies an old idea about theory change known as the Duhem-Quine thesis (Duhem, 1991). The idea can illustrated by a simile, originally attributed to Otto Van Neurath (1932/1983) but popularized by Quine, who writes:

> "*We* [theorists] *are like sailors who on the open sea must reconstruct their ship but are never able to start afresh from the bottom. Where a beam is taken away a new one must at once be put there, and for this the rest of the ship is used as support. In this way, by using the old beams and driftwood the ship can be shaped entirely anew, but only by gradual reconstruction.*" (1969, p3)

The *Neurath's ship* metaphor describes the piecemeal growth and evolution of scientific theories over the course of history. In the metaphor, the theorist (sailor) is cast as relying on their existing theory (ship) to stay afloat, without the privilege of a dry-dock in which to make major improvements. Unable to step back and consider all possible alternatives, the theorist is limited to building on the existing theory, making a series of small changes with the goal of improving the fit.

We argue that people are in a similar position when it comes to their beliefs about the causal structure of the world. We propose that a learner normally maintains only a single hypothesis about the global causal model, rather than a distribution over all possibilities. They update their hypothesis by making local changes (e.g. adding, removing and reversing individual connections, nodes or subgraphs) while depending on the rest of the model as a basis. We show that by doing this, the learner can end up with a relatively accurate causal model without ever representing the whole hypothesis space or storing all the old evidence, but that their causal beliefs will exhibit a particular pattern of sequential dependence. We provide a related account of bounded intervention selection, based on the idea that learners adapt to their own learning limitations when choosing what evidence to gather next, attempting to resolve local rather than global uncertainty. Together, our *Neurath's ship* model and local-uncertainty-based schema for intervention selection provide a step towards an explanation of how people might achieve a *resource rational* (Simon, 1982; **?**) trade-off between accuracy and the cognitive costs

of maintaining an accurate causal model of the world.

The paper is organized as follows. We first formalize causal model inference at the computational level. We then highlight the ways in which past experiments have shown human learning to diverge from the predictions of this idealized account, using these to motivate two causal judgment heuristics proposed in the literature: *simple endorsement* (Bramley, Lagnado, & Speekenbrink, 2015; Fernbach & Sloman, 2009) and *win-stay, lose-sample* (Bonawitz, Denison, Gopnik, & Griffiths, 2014) before developing out own *Neurath's ship* framework for belief change and active learning.

We next show that participants' overall patterns of judgments and intervention choices are in line with the predictions of our framework across a variety of problems varying in terms of the complexity and noise in the true generative model, and whether the participants' are trained or must infer the noise.

We then compare models at the individual level, showing that all three causal-judgment proposals substantially outperform baseline and computational level competitors. While our Neurath's ship provides the best overall fit, we find considerable diversity of strategies across participants. In particular, we find that the *simple endorsement* heuristic emerges as a strong competitor. We provide additional details about the formal framework and model specification in the Appendix. Also, where indicated, additional figures are provided in Supplementary materials at `http://www.ucl.ac.uk/lagnado-lab/el/ns_sup`.

A computational-level framework for active structure learning

Before presenting our theoretical framework, we lay out our computational-level analysis of the problem of structure learning. This can be broken down into three interrelated elements: (1) representing causal models (2) performing inference over possible models given evidence (observations and the outcomes of interventions), and (3) selecting interventions to gather more evidence and support this inference. We introduce the three elements here, providing more detail where indicated in Appendix A.

Representation

We use a standard representation for causal models – the parameterized directed acyclic graph (Pearl, 2000, see Figure 1a). Nodes represent variables (i.e. the component parts of a causal system); arrows represent causal connections; and parameters encode the combined influence of parents (the source of an arrow) on children (the arrow's target)[1]. Such graphs can represent continuous variables and any form of causal relationship; but here we focus on systems of binary $\{0 = \text{absent}, 1 = \text{present}\}$ variables and and assume generative connections – meaning we assume that the presence of a cause will always *raise* the probability that the effect is also present.[2]

[1]Following standard graph nomenclature, we will often refer to the space between a pair of nodes in a model as an "edge", so that an acyclic causal model defines each edge as taking one of three states: forward $\rightarrow$, backward $\leftarrow$, or inactive $\varnothing$.

[2]It is worth noting that these graphs cannot naturally represent cyclic or reciprocal relationships. However, there are various ways to extend the formalism as we discuss in General Discussion, and our theory is not fundamentally

We also adopt Cheng's Power PC (1997) convention for parameterization, which provides a simple way to capture how probabilistic causal influences combine. This assumes that causes have independent chances of producing their effects, meaning the probability that a variable takes the value 1 is a noisy-OR combination of the *power* or strength w_S of any active causes of it in the model (we assume that this value is the same for all connections), together with that of an omnipresent background strength w_B encapsulating the influence of any causes exogenous to the model. We write $\mathbf{w} = \{w_S, w_B\}$. The probability that variable x takes the value 1 is thus

$$P(x = 1|pa(x), \mathbf{w}) = 1 - (1 - w_B)(1 - w_S)^{\sum_{y \in pa(x)} y} \quad (1)$$

where $pa(x)$ denotes the parents of variable x in the causal model (see Figure 1a for an example). For convenience, we assume $\mathbf{w}$ is the same for all components.[3]

Inference

Each causal model m over variables X with strength and background parameters $\mathbf{w}$, assigns a probability to each datum $\mathbf{d} = \{x \dots z\}$, propagating information from the variables that are fixed through intervention $\mathbf{c}$, to the others (see Figure 1b). The space of all possible interventions $\mathcal{C}$ is made up of all possible combinations of fixed and unfixed variables, and for each intervention $\mathbf{c}$ the possible data $\mathcal{D}_\mathbf{c}$ is made up of all combinations of absent/present on the unfixed variables. We use Pearl's Do[.] operator (Pearl, 2000) to denote what is fixed on a given test. For instance, $\mathrm{Do}[x = 1, y = 0]$ means a variable x has been fixed "on" and variable y has been fixed "off", with all other variables free to vary[4]. Interventions allow a learner to override the normal flow of causal influence in a system, initiating activity at some components and blocking potential influences between others. This means they can provide information about the presence and direction of influences between variables that is typically unavailable from purely observational data (see Bramley, Lagnado, & Speekenbrink, 2015; Pearl, 2000, for a more detailed introduction), without additional cues such as temporal information (Bramley, Gerstenberg, & Lagnado, 2014). For instance, in Figure 1b, we fix y to 1 and leave x and z free ($\mathbf{c} = \mathrm{Do}[y = 1]$). Under the $x \rightarrow y \rightarrow z$ model we would then expect x to activate with probability w_B and z with a probability of $1 - (1 - w_B)(1 - w_S)$.

In total, the probability of datum $\mathbf{d}$, given intervention $\mathbf{c}$, is just the product of the probability of each variable that was not intervened upon, given the states of its parents in the model

$$P(\mathbf{d}|m, \mathbf{w}, \mathbf{c}) = \prod_{x \in (X \notin \mathbf{c})} P(x|\{\mathbf{d}, \mathbf{c}\}_{pa(x)}, \mathbf{w}). \quad (2)$$

where $\{\mathbf{d}, \mathbf{c}\}_{pa(x)}$ indicates that those parents might either be observed (part of $\mathbf{d}$) or fixed by the intervention (part of $\mathbf{c}$).

In fully Bayesian inference, the true model is considered to be a random variable M. Our

tied to a particular representation.

[3]We also restrict ourselves to cases without any latent variable, although we note that imputing the presence of hidden variables is another important and computationally-challenging component of causal inference (Buchanan, Tenenbaum, & Sobel, 2010; Kushnir, Gopnik, Lucas, & Schulz, 2010).

[4]We include the pure observation $Do[\varnothing]$ in $\mathcal{C}$.

prior belief $P(M)$ is then an assignment of probabilities, adding up to 1 across possible models $m \in M$ in the set of models $\mathcal{M}$. When we observe some data $D = \{\mathbf{d}^i\}$, associated with interventions $C = \{\mathbf{c}^i\}$, we can update these beliefs with Bayes theorem by multiplying our prior by the probability of the observed data under each model and dividing by the weighted average probability of those data across all the possible models:

$$P(m|D, \mathbf{w}; C) = \frac{P(D|m, \mathbf{w}; C)P(m)}{\sum_{m' \in M} P(D|m', \mathbf{w}; C)P(m')}. \tag{3}$$

We will typically treat the data as being independent and identically distributed, so $P(D|m, \mathbf{w}; C) = \prod_i P(\mathbf{d}^i|m, \mathbf{w}; \mathbf{c}^i)$.

If the data arrive sequentially (as $D^t = \{\mathbf{d}^1, \ldots, \mathbf{d}^t\}$; and similarly for the interventions), we can either store them and update at the end, or update our beliefs sequentially, taking the posterior $P(M|D^{t-1}, \mathbf{w}; C^{t-1})$ at timestep $t-1$ as the new "prior" for datum $\mathbf{d}^t$. If we are also unsure about the parameters of the true model (i.e. w_B and w_S) we have to treat them as random variables too and average over our uncertainty about them to compute a marginal posterior over models M (see Appendix A).

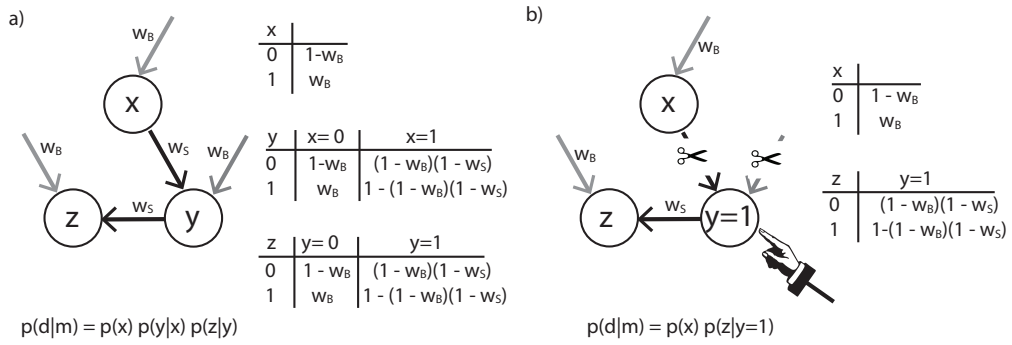

Figure 1: Causal model representation. a) An example causal Bayesian network, parametrized with strength w_S and base rate w_B. The tables give the probability of each variable taking the value 1 conditional on its parents in the model and the omnipresent background noise rate w_B. b) Visualization of intervention $\mathrm{Do}[y = 1]$. Setting y to 1 renders it independent of its normal causes as indicated by the scissors symbols.

Choosing interventions

It is clear that different interventions yield different outcomes, which in turn have different probabilities under different models. This means that which interventions are valuable for identifying the true model depends strongly on the hypothesis space and prior. For instance fixing y to 1 ($\mathrm{Do}[y = 1]$) is (probabilistically) diagnostic if you are primarily unsure whether x causes z because $p(z|\,\mathrm{Do}[y = 1])$ differs depending whether $pa(z)$ includes x. However, it is not diagnostic if you are primarily unsure whether x causes y because y will take the value 1 the same regardless of whether $\mathrm{pa}(x)$ includes y.

The value of an intervention can be quantified relative to a notion of uncertainty. We can define the value of an intervention as the expected reduction in uncertainty about the true model after seeing its outcome.[5] To calculate this expectation, we must average, prospectively,

[5]Strictly this is greedy rather than optimal because planning several steps ahead can result in a different intervention

over the different possible outcomes $\mathbf{d}' \in \mathcal{D}_{\mathbf{c}}$ (recalling $\mathcal{D}_{\mathbf{c}}$ is the space of possible outcomes of intervention $\mathbf{c}$) weighted by their marginal likelihoods under the prior. For a greedily optimal sequence of interventions $\mathbf{c}^1, \ldots, \mathbf{c}^t$, we take $P(M|D^{t-1}, \mathbf{w}; C^{t-1})$ as our prior each time. The most valuable intervention $\mathbf{c}^t$ at a given time point is then

$$\arg\max_{\mathbf{c}\in\mathcal{C}} \mathop{\mathbb{E}}_{\mathbf{d}'\in\mathcal{D}_{\mathbf{c}}} \left[\Delta H(M|\mathbf{d}', D^{t-1}, \mathbf{w}; C^{t-1}, \mathbf{c})\right], \tag{4}$$

where $\mathbb{E}[.]_{\mathbf{d}'\in\mathcal{D}_{\mathbf{c}}}$ denotes the average over outcomes $\mathbf{d}'$ and $\Delta H(.)$ denotes reduction in uncertainty. We use Shannon entropy (**?**) to measure uncertainty (see Appendix A). Shannon is just one of a broad family of possible entropy measures (Nielsen & Nock, 2011). However, it is one that has proved at least as long-run successful as a number of variants when applied as a greedy strategy for choosing interventions (Bramley, Nelson, Speekenbrink, & Lagnado, 2014) or asking binary questions (Nelson, 2005).

Behavioral patterns and existing explanations

Unfortunately, both inference and choosing interventions scale so poorly in the number of variables, they are fundamentally intractable for any plausibly bounded learner (Cooper, 1990; van Rooij, Wright, Kwisthout, & Wareham, 2014). The number of possible graphs grows rapidly with the number of variables they relate (3-, 4- and 5-variable problems have 25, 543 and 29281 respectively). Active intervention selection adds extra complexity because there are many possible interventions (3-,4- and 5- variable problems permit 27, 81 and 243 patterns of fixed "on", fixed "off" and free components), each of which might yield many outcomes (up to 8, 16 and 32 respectively, depending how many variables are left free to vary). All combinations of potential model, intervention and outcome should be averaged over in order to select the most valuable intervention. This implies that people must find a considerably more economical way to approximate model inference while maintaining satisfactory accuracy.

It is therefore not surprising that behavioral learning patterns in existing studies exhibit marked divergence from the predictions of idealized Bayesian learning. Participants' model judgments are typically robustly better than chance, yet poor when compared directly against an idealized Bayesian learner (Bramley, Gerstenberg, & Lagnado, 2014; Fernbach & Sloman, 2009; Lagnado & Sloman, 2002, 2004). Likewise, adults and even children have been shown to select interventions that are robustly more informative than chance, but much less efficient than idealized active learning (Bramley, Lagnado, & Speekenbrink, 2015; Coenen et al., 2015; Gureckis & Markant, 2009; Lucas, Bridgers, Griffiths, & Gopnik, 2014; Markant & Gureckis, 2012; McCormack, Bramley, Frosch, Patrick, & Lagnado, 2016; Steyvers et al., 2003).

More revealing than mere performance are the ways in which participants' judgments diverge from these rational norms (Anderson, 1990). Bramley, Lagnado, and Speekenbrink (2015) found that participants' judgments in a sequential active causal learning task resembled probability matching when lumped together, but that individuals' trajectories were not well captured by simply adding decision noise to the Bayesian predictions. Individuals' sequences

being favored. However, planning ahead was shown to make little difference for the similar problems explored in Bramley et al (2015).

of judgments were much too sequentially dependent, or "sticky", compared to the Bayesian predictions, tending to remain the same or similar over multiple elicitations as the objectively most likely structure shifted. At the same time, when participants did change their judgments, they tended to do so in ways that were consistent with the most recently gathered evidence, neglecting evidence gathered earlier in learning. The result was a dual pattern of recency in terms of judgments' consistency with the evidence, and stickiness in terms of consistency with the previous judgments. Bramley et al found that they could capture these patterns with the addition of two parameters to the Bayesian model. The first was a forgetting parameter, encoding trial-by-trial leakage of information from the posterior as it became the prior for the next test. The second was a conservatism parameter, encoding a non-normatively high probability assigned to the latest causal hypothesis. While the resulting model captured participants choices, it still made the implausible assumption that learners maintained weighted probabilistic beliefs across the whole hypothesis space and performed efficient active learning with respect to these.

As with Bramley et al, Bonawitz et al. (2014) found that children and adults' online structure judgments exhibited sequential dependence. To account for this they proposed an account of how causal learners might rationally reduce the computational effort of continually reconsidering their model. In their "win-stay, lose-sample" scheme they suggest that learners maintain a single structural hypothesis, only resampling a new hypothesis from the posterior when they see something surprising under their current model, concretely, with a probability that increases as the most recent observation becomes less probable. This scheme guarantees that the learner's latest hypothesis is a sample from the posterior distribution at every point, but does not the require them to resample with every new trial. While it captures the intuitive idea that people will tend to stick with a hypothesis until it fails to perform, "win-stay, lose-sample" still requires the learner to store all the past evidence to use when resampling, and does not provide a recipe for how the samples are drawn.[6]

Another approach to understanding deviations between people's causal judgments and rational norms comes from the idea that people construct causal models in a modular or piecewise way. For example, Waldmann, Cheng, Hagmayer, and Blaisdell (2008) propose a *minimal rational model* under which learners infer the relationships between each pair of variables separately without worrying about the dependencies between them, ending up with a modular causal model that allows for good local inferences but which leads to so-called "Markov violations" in more complex inferences where participants fail to respect the conditional dependencies and independences implied by the global model (Rehder, 2014). They show that this minimal model is sufficient to capture participants judgment patterns in two case studies. Building on this idea of locality, Fernbach and Sloman (2009) asked participants to make judgments following observation of several preselected interventions. They found that participants were particularly bad at inferring chains, often inferring spurious additional links from the root to the sink node (e.g. $x \rightarrow z$ as well as $x \rightarrow y$ and $y \rightarrow z$), a pattern also observed in Bramley, Lagnado, and Speekenbrink (2015). Fernbach and Sloman proposed that this was

[6]The authors mention that MCMC could be used to draw these samples without representing the full posterior.

a consequence of participants inferring causal relationships through local rather than global computations. In the example, the interventions on x would normally lead to activations of z due to the indirect connection via y. If learners attended only to x and z there would be the appearance of a direct relationship. They found that they could better model participants by assuming they inferred each causal link separately while ignoring the the rest of the model. Embodying this principle, (Bramley, Lagnado, & Speekenbrink, 2015) proposed a *simple endorsement* heuristic for online causal learning that would tend to add direct edges to a model between intervened-on variables and any variables that activated as a result, removing edges going to any variables that didn't activate. By doing this after each new piece of evidence, the model exhibited recency as the older edges would tend to be overwritten by newly inferred ones, as well as as capturing the pattern of adding unnecessary direct connections in causal chains. The model did a good job of predicting participants' patterns but was outperformed by the Bayesian model bounded with forgetting and conservatism. Additionally, like any heuristic, *simple endorsement*'s success is conditional on its match to the situation. For instance, *simple endorsement* does badly in cases where there are many chains – meaning that the outcome of many interventions are indirect, and also if the true w_B is high.

Going beyond causal learning, sequential effects are ubiquitous in cognition. In some instances they can be rational; for instance moderate recency is rational in a changing world (Julier & Uhlmann, 1997). Regardless, there are a plethora of non Bayesian models that can reproduce various sequential effects (DeCarlo, 1992; Gilden, 2001; Treisman & Williams, 1984). A common class of these is based on the idea of adjusting an estimate part way toward new evidence (e.g. Einhorn & Hogarth, 1986; Petrov & Anderson, 2005; Rescorla, Wagner, et al., 1972). Updating point estimates means that a learner need not keep all the evidence in memory but can instead make use of the location of the point(s) as a proxy for what was learned in the past. Bramley, Dayan, and Lagnado (2015) propose a model inspired by these ideas, that maintains a single hypothesis, but simultaneously attempts to minimize edits along with the number of variables' latest states that the current model fails to explain. The result is a model where the current belief acts as an anchor and the learner tends to try to explain the latest evidence by making the minimal number of changes to it. Again, this model provided a good fit with participants' judgments, but did not provide a procedure for how participants were able to search the hypothesis space for the causal structure that minimized these constraints.

In summary, a number of ideas and models have been proposed in the causal and active learning literatures. By design, they all do a good job of capturing patterns in human causal judgments. However, it is not clear that any of these proposals provide a general purpose, scalable explanation for human success in learning a complex causal world-model. Some (e.g *win-stay, lose-sample*) capture behavioral patterns within the normative framework, but do not provide a scalable algorithm. Others (e.g. *simple endorsement*) provide simple scalable heuristics but may not generalize beyond the tasks they were designed for, nor explain human successes in harder problems. In the next section we take inspiration from methods for approximate inference in machine learning to construct a general purpose algorithm for incremental structure change that satisfies both these desiderata.

Algorithms for causal learning with limited resources

We now turn to algorithms in machine learning that make approximate learning efficient in otherwise intractable circumstances. Additionally, research in these fields on active learning and optimal experiment design has identified a range of reasonable heuristics for selecting queries when the full expected information calculation of (Equation 4) is intractable. We will take inspiration from some of these ideas to give a formal basis to the intuitions behind the *Neurath's ship* metaphor. We will then use this formal model to generate predictions that we will compare to participants' behavior in several experiments.

Approximating with a few hypotheses

One common approximation, for situations where a posterior cannot be evaluated in closed form, is to maintain a manageable number of individual hypotheses, or "particles" (Liu & Chen, 1998), with weights corresponding to their relative likelihoods. The ensemble of particles then acts as an approximation to the desired distribution. Sophisticated reweighting and resampling schemes can then filter the ensemble as data are observed, approximating Bayesian inference.

These "particle filtering" methods have been used to explain how humans and other animals might approximate the solutions to complex problems of probabilistic inference. In associative learning (Courville & Daw, 2007), categorization (Sanborn, Griffiths, & Navarro, 2010) and binary decision making (Vul, Goodman, Griffiths, & Tenenbaum, 2009), it has been proposed that people's beliefs actually behave most like a single particle, capturing why individuals often exhibit fluctuating and sub-optimal judgment while maintaining a connection to Bayesian inference, particularly at the population level.

Sequential local search

The idea that people's causal theories are like particles requires they also have some procedure for sampling or adapting these theories as evidence is observed. Another class of useful machine learning methods involves generating sequences of hypotheses, each linked to the next via a form of possibly stochastic transition mechanism. Two members of this class are particularly popular in the present context: Markov Chain Monte Carlo (MCMC) sampling, which asymptotically approximates the posterior distribution; and (stochastic) hill climbing, which merely tries to find hypotheses that have high posterior probabilities.

MCMC algorithms involve stochastic transitions with samples that are typically easy to generate. Under various conditions, this implies that the sequences of (dependent) sample hypotheses form a Markov chain with a stationary distribution that is the full, intended, posterior distribution (Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953). The samples will appear to "walk" randomly around space of possibilities, tending to visit more probable hypotheses more frequently. If samples are extracted from the sequence after a sufficiently long initial, so-called burn-in, period, and sufficiently far apart (to reduce the effect of dependence), they can provide a good approximation to the true posterior distribution. There are typically many different classes of Markov chain transitions that share the same stationary distribution,

but differ in the properties of burn-in and subsampling.

The stochasticity inherent in MCMC algorithms implies that the sequence sometimes makes a transition from a more probable to a less probable hypothesis – this is necessary to sample multi-modal posterior distributions. A more radical heuristic is only to allow transitions to more probable hypotheses — this is called "hill-climbing", attempting to find, and then stick at, the best hypothesis (Tsamardinos, Brown, & Aliferis, 2006). This is typically faster than a full MCMC algorithm to find a good hypothesis, but is prone to become stuck in a local optimum, where the current hypothesis is more likely than all its neighbors, but less likely than some other more distant hypothesis.

Applied to causal structure inference, we might in either case consider transitions that change at most a single edge in the model (Cooper & Herskovits, 1992; Goudie & Mukherjee, 2011). A simple case is Gibbs sampling (Geman & Geman, 1984), starting with some structural hypothesis and repeatedly selecting an edge (randomly or systematically) and re-sampling it (either adding, removing or reversing) conditional on state of the other edges. This means that a learner can search for a new hypothesis by making local changes to their current hypothesis, reconsidering each of the edges in turn, conditioning on the state of the others without ever enumerating all the possibilities. By constructing a short chain of such "rethinks" a learner can easily update a singular hypothesis without starting from scratch. The longer the chain, the less dependent or "local" the new hypothesis will be to the starting point.

The idea that stochastic local search plays an important role in cognition has some precedent (Gershman, Vul, & Tenenbaum, 2012; Sanborn et al., 2010). For instance, Abbott, Austerweil, and Griffiths (2012) propose a random local search model of memory retrieval and Ullman, Goodman, and Tenenbaum (2012) propose an MCMC search model for capturing how children search large combinatorial theory spaces when learning intuitive physical theories like taxonomy and magnetism. The idea that people might update their judgments by something like MCMC sampling is also explored by Lieder, Griffiths and Goodman (2012; under review). They argue that under reasonable assumptions about the costs of resampling and need for accuracy, it can be rational to update one's beliefs by constructing short chains where the the updated judgment retains some dependence on its starting state, arguing that this might explain *anchoring* effects (Kahneman, Slovic, & Tversky, 1982).

In addition to computational savings, updating beliefs by local search can be desirable for statistical reasons. If the learner has forgotten some of the evidence they have seen, the location of their previous hypothesis acts like a very approximate version of a sufficient statistic for the forgotten information. This can make it advantageous to the learner to strike a good balance between editing their model to account better for the data they can remember, and staying close to their previous model to retain the connection to the data they have forgotten (Bramley, Lagnado, & Speekenbrink, 2015).

Neurath's ship: An algorithmic-level model of sequential belief change

The previous section summarized two ideas derived from computer science and statistics that provide a potential solution to the computational challenges of causal learning: maintaining only a single hypothesis at a time, and exploring new hypotheses using local search based on sampling. In this section, we formalize these ideas to define a class of models of causal learning inspired by the metaphor of Neurath's ship. We start by treating interventions as given, and only focus on inference. We then consider the nature of the interventions.

Concretely, we propose that causal learners maintaining only a single causal model (a single particle), b^{t-1} and a collection of recent evidence and interventions $\mathcal{D}_r^{t-1}$ and $\mathcal{C}_r^{t-1}$ at time $t-1$. They then make inferences by:

1. Observing the latest evidence $\mathbf{d}^t$ and $\mathbf{c}^t$ and adding it to the collection to make $\mathcal{D}_r^t$ and $\mathcal{C}_r^t$.
2. Then, searching for local improvements to b^{t-1} by sequentially reconsidering edges $E_{ij} \in \{1 : i \rightarrow j,\ 0 : i \nleftrightarrow j,\ -1 : i \leftarrow j\}$ (adding, subtracting or reorienting them) conditional on the current state of the edges in the rest of their model $E_{\setminus ij}$ – e.g. with probability $P(E_{ij}|E_{\setminus ij}, \mathcal{D}_r^t, \mathcal{C}_r^t, \mathbf{w})$.
3. After searching for k steps, stopping and taking the latest version of their model as their new belief b^t. If b^t differs from b^{t-1} the evidence is forgotten ($\mathcal{D}_r^t$ and $\mathcal{C}_r^t$ become $\{\}$), and they begin collecting evidence again.

A detailed specification of this process is given in Appendix A.

Starting with any hypothesis and repeatedly resampling edges conditional on the others is a form of Gibbs sampling (Goudie & Mukherjee, 2011). Further, the learner can make use of the data they have forgotten by starting the search with their current belief b^{t-1}, since these data are represented to some degree in the location of b^{t-1}. Resampling using the recent data $P(M|\mathcal{D}_r^t, \mathcal{C}_r^t, \mathbf{w})$ allows the learner to adjust their beliefs to encapsulate better the data they have just seen, and let this evidence fall out of memory once it has been incorporated into the model.

Resampling, hill climbing or random change

Following the procedure outlined above, the learner's search steps would constitute dependent samples from the posterior over structures given $\mathcal{D}_r^t$. However, it is also plausible that learners will try to hill-climb rather than sample, preferring to move to more probable local models more strongly than would be predicted by Gibbs sampling. In order to explore this, we will consider generalizations of of the update equation allowing transitions to be governed by powers of the conditional edge probability (i.e. $P^\omega(E_{ij} = e|E_{\setminus ij}, \mathcal{D}_r^t, \mathcal{C}_r^t, \mathbf{w})$), yielding stronger or weaker preference for the most likely state of E_{ij} depending whether $\omega > 1$ or < 1. By setting ω to zero, we would get a model that does not learn but just moves randomly between hypotheses, tending to remain local and by setting it to infinity we would get a model that always moved to the most likely state for the edge.

Search length

It is reasonable to assume that the number of search steps k that a learner performs will be variable, but that their *capacity* to search will be relatively stable. Therefore, we assume that for each update, the learner searches for k steps, where k is drawn from a Poisson distribution with mean $\lambda \in [0, \infty]$.

The value of λ thus determines how sequentially dependent a learner's sequences of beliefs are. A large λ codifies a tendency to move beliefs a long way to account for the latest data $\mathcal{D}_r^t$ at the expense of the older data – retained only in the location of the previous belief b^{t-1} – while a moderate λ captures a reasonable trade-off between starting state and new evidence, and a small λ captures *conservatism*, i.e. failure to shift beliefs enough to account for the latest data.[7]

Putting these together

By representing the transition probabilities from model i to model j, for a particular setting of hill climbing parameter ω and data $\mathcal{D}_r^t$, with a transition matrix R_t^ω, we can thus make probabilistic predictions about a learner's new belief $b^t \in B^t$.[8] The probabilities depend on the previous belief b^{t-1} and their average search length λ. By averaging over different search lengths with their probability controlled by λ, and taking the requisite row of the resulting transition matrix we get the following equation

$$P(b^t = m|\mathcal{D}_r^t, \mathcal{C}_r^t, b^{t-1}, \omega, \lambda) = \sum_0^\infty \frac{\lambda^k e^{-\lambda}}{k!} [(R_t^\omega)^k]_{b^{t-1}m} \tag{5}$$

Note that this equation describes the probability of a Neurath's ship style search terminating in a given new location. The learner themselves need only follow the four steps described above, *sampling* particular edges and search length rather than averaging over the possible values of these quantities. See Appendix A for more details and Figure 2 for an example.

[7] Note that we later cap k at 50 when estimating our model having established that search lengths beyond these bounds made negligible difference to predictions.

[8] We define this matrix formally in Appendix A. Note that we assume transitions that would create a loop in the overall model get a probability of zero. This assumption could be dropped for learning dynamic Bayesian networks but is necessary for working with directed acyclic graphs.

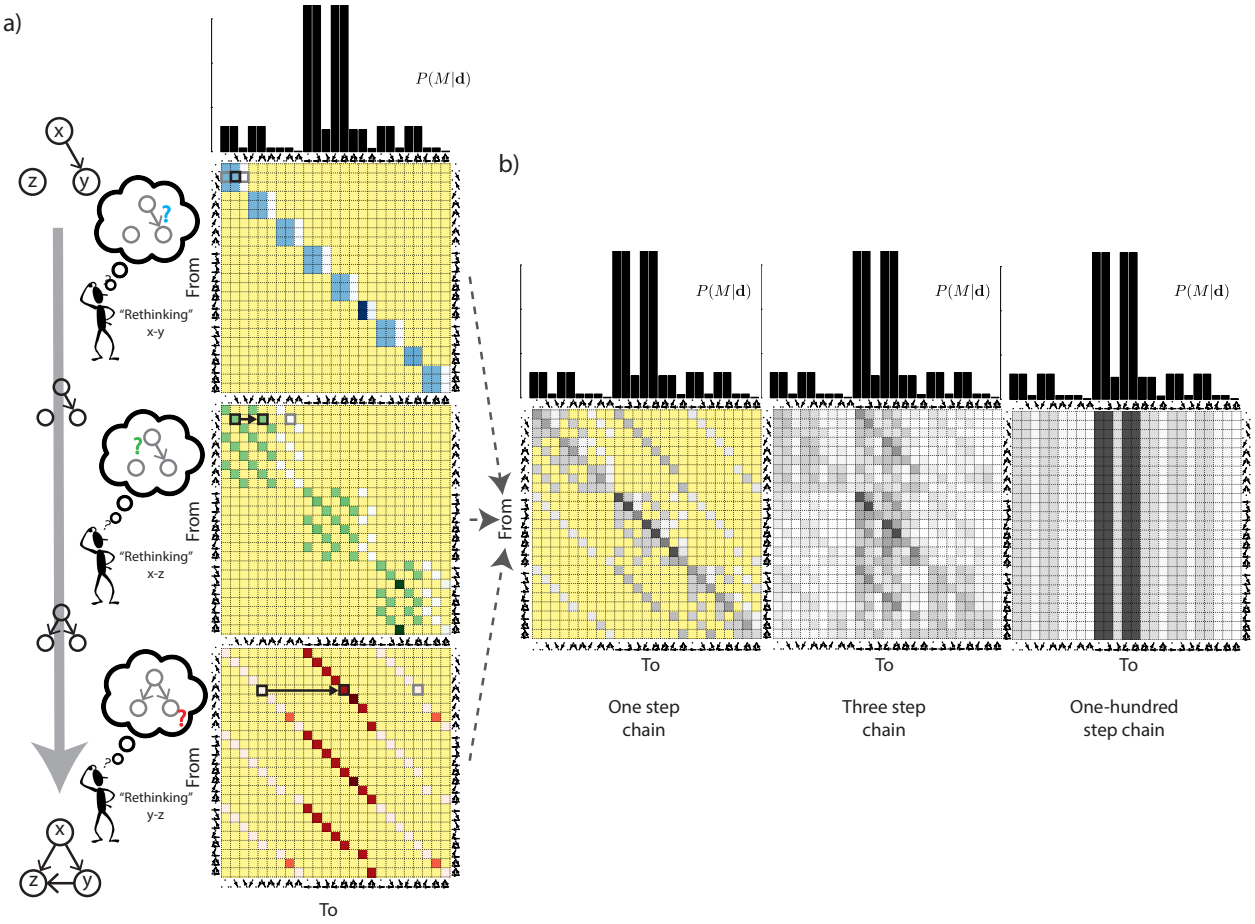

Figure 2: An illustration of NS model of causal belief updating. a) An example search path: The learner starts out with a singly connected model at the top ($x \rightarrow y$ connection only). They update their beliefs by resampling one edge at a time $e \in \{\rightarrow, \nleftrightarrow, \leftarrow\}$. Each entry i, j in the matrices gives the probability of moving from model in the row i to the model in the column j when resampling the edge marked with the colored question mark. Lighter shades of the requisite color indicate low transition probability, darker shades indicate greater transition probability; yellow is used to indicate zero probabilities. Here the learner stops after resampling each edge once, moving from b^{t-1} of $[x \rightarrow y]$ to b^t of $[x \rightarrow y, x \rightarrow z, y \rightarrow z]$. b) Assuming the edge to resample is chosen at random, we can average over the different possible edge choices to derive a 1-step Markov chain transition matrix R_t^ω encompassing all the possibilities. By raising this matrix to higher powers we get the probability of different end points for searches of that length. If the chain is short (small k) the final state depends heavily on the starting state (left) but for longer chains (large k), the starting state becomes less important, getting increasingly close to independent sampling from the desired distribution (right).

Selecting interventions on Neurath's ship: A local uncertainty schema

In situations where a posterior is already hard to evaluate, calculating the globally most informative intervention – finding the intervention $\mathbf{c}^t$ that maximizes Equation 4 – will almost always be infeasible. Therefore, a variety of heuristics have been developed that allow tests to be selected that are more useful than random selection, but do not require the full expected information gain be computed (Settles, 2012). These tend to rely on the learners' current, rather than expected, uncertainty (e.g. *uncertainty sampling* which chooses based on outcome uncer-

tainty under the prior) or the predictions under just a few favored hypotheses (e.g. *query by committee*) as a substitute for the full expectancy calculation. The former relies on maintaining a complete prior distribution, making the latter a more natural partner to the *Neurath's ship* framework.

We have proposed a model of structure inference under which learners are only able to consider a small set of of alternatives at a time, and only able to generate alternatives that are "local" in some dimension. Locally driven intervention selection is a natural partner to this for at least two reasons: (1) Under the constraints of the *Neurath's ship* framework, learners would not be able to work with the prospective distributions required to estimate global expected informativeness, but could potentially estimate expected informativeness with respect to a sufficiently narrow sets of alternatives. (2) Evidence optimized to distinguishing local possibilities (focused on one edge at a time for instance) might better support sequential local belief updates (of the kind emphasized in our framework) than the globally most informative evidence (Patil, Zhu, Kopeć, & Love, 2014). In line with this, we propose one way in which learners might select robustly informative interventions by attempting only to distinguish a few "local" possibilities at a time, requiring only "local" uncertainty estimates to target the possibilities on which to focus (Markant, Settles, & Gureckis, 2015).

The idea that learners will focus on distinguishing only a few alternatives at a time requires specifying how they choose which of the many possible subsets of the full hypothesis space to target with a particular test. Queries that optimally reduce expected uncertainty about one local aspect of a problem are liable to differ from those that promise high global uncertainty reduction. For example, Figure 3b shows two trials taken from our experiments, and shows that the expected values of each of a range of different intervention choices (shown in Figure 3a) are very different depending whether the learner is focused on resolving global uncertainty all at once, or on resolving some specific "local" aspect of it. This illustrates the idea that a learner might choose a test that is optimally informative with respect to a modest range of options that they have in mind at the time (e.g. models that differ just in terms of the state of E_{xz}) yet appear sporadically inefficient from the perspective of greedy global uncertainty reduction. Furthermore, by licensing quite different intervention preferences, they allow us to diagnose individual and trial-by-trial differences in focus preference.

In the current work, we will consider three possible varieties of focus, one motivated by the *Neurath's ship* framework (*edge* focus) and two inspired by existing ideas about bounded search and discovery in the literature (*effects* focus and *confirmation* focus). While these are by no means exhaustive they represent a reasonable starting point.

The two stages of the schema

The idea that learners focus on resolving local rather than global uncertainty results in a metaproblem of choosing what to focus on next, making intervention choice a two stage process. We write L for the set of all possible foci $\mathbf{l}$, and $\mathcal{L} \subset L$ for the subset of possibilities that the learner will consider at a time, such as the the state of a particular edge or the effects of a particular variable. The procedure is:

Stage 1 Selecting a local focus $\mathbf{l}^t \in \mathcal{L}$

Stage 2 Selecting an informative test $\mathbf{c}^t$ with respect to the chosen focus $\mathbf{l}^t$

Different learners might differ in the types of questions they consider, meaning that $\mathcal{L}$ might contain different varieties and combinations of local focuses. We first formalize the two stages of the schema, and then propose three varieties of local focus that learners might consider in their option set $\mathcal{L}$ that differ in terms of which and how many alternatives they include.

As mentioned above, we assume that the learner has some way of estimating their current local confidence. We will assume confidence here is approximately the inverse of uncertainty, so assume for simplicity that learners can calculate uncertainty from the evidence they have gathered since last changing their model in the form of the entropy $H(\mathbf{l}|\mathcal{D}_r^t, \mathbf{w}; \mathcal{C}_r^t)$ for all $\mathbf{l} \in \mathcal{L}$ (the assumption we examine in the discussion). They then choose (**Stage 1**) the locale where these data imply the least certainty

$$\mathbf{l}^t = \arg\max_{\mathbf{l}\in\mathcal{L}} H(\mathbf{l}|b^{t-1}, \mathcal{D}_r^t, \mathbf{w}; \mathcal{C}_r^t) \tag{6}$$

However, in carrying out **Stage 2** we make the radical assumption that learners do not use $P(\mathbf{l}^t|\mathcal{D}_r^t, b^{t-1}, \mathbf{w}; \mathcal{C}_r^t)$, but rather, consistent with the method of inference itself, only consider the potential next datum $\mathbf{d}'$. This means that the intervention $\mathbf{c}^t$ itself is chosen to maximize the expected information about $\mathbf{l}^t$, ignoring pre-existing evidence, and using what amounts to a uniform prior. Specifically, we assume that $\mathbf{c}^t$ is chosen as

$$\mathbf{c}^t = \arg\max_{\mathbf{c}\in\mathcal{C}} \mathop{\mathbb{E}}_{\mathbf{d}\in\mathcal{D}_\mathbf{c}} \left[\Delta H(\mathbf{l}^t|\mathbf{d}, \mathbf{w}, b^{t-1}; \mathbf{c})\right] \tag{7}$$

where we detail the term in the expectation below for the three types of focuses.

Assuming real learners will exhibit some decision noise, we can model both choice of focus and choice of intervention relative to a focus as soft (Luce, 1959) rather than strict maximization giving focus probabilities

$$P(\mathbf{l}^t|\mathcal{D}_r^t, b^{t-1}, \mathbf{w}; \mathcal{C}_r^t) = \frac{\exp(H(\mathbf{l}^t|\mathcal{D}_r^t, b^{t-1}, \mathbf{w}; \mathcal{C}_r^t)\rho)}{\sum_{\mathbf{l}\in\mathcal{L}} \exp(H(\mathbf{l}|\mathcal{D}_r^t, b^{t-1}, \mathbf{w}; \mathcal{C}_r^t)\rho)} \tag{8}$$

governed by some inverse temperature parameter ρ, and choice probabilities

$$P(\mathbf{c}^t|\mathbf{l}, \mathbf{w}, b^{t-1}) = \frac{\exp(\mathbb{E}_{\mathbf{d}'\in\mathcal{D}_\mathbf{c}} \left[\Delta H(\mathbf{l}|\mathbf{d}', \mathbf{w}, b^{t-1}; \mathbf{c}^t)\right] \eta)}{\sum_{c\in\mathcal{C}} \exp(\mathbb{E}_{\mathbf{d}'\in\mathcal{D}_\mathbf{c}} \left[\Delta H(\mathbf{l}|\mathbf{d}', \mathbf{w}, b^{t-1}; \mathbf{c})\right] \eta)} \tag{9}$$

governed by inverse temperature η.

Three varieties of local focus

Edges

An obvious choice, given the *Neurath's ship* framework, would be for learners to try to distinguish alternatives that differ in terms of a single edge (Figure 3a), i.e. those they would consider during a single update step.

For a chosen edge E_{xy} we can then consider a learner's goal to be to maximise their expec-

tation of

$$\Delta H(E_{xy}|E^{t-1}_{\setminus xy}, \mathbf{d}, \mathbf{w}; \mathbf{c}) \tag{10}$$

(see Appendix A for the full local entropy equations). Note that Equation 10 is a refinement of Equation 7 for the case of focusing on an edge, from b^{t-1} the learner need only condition on the other edges $E^{t-1}_{\setminus xy}$. This goal results in a preference for fixing one of the nodes of the target edge "on", leaving the other free, and depending on the other connections in b^{t-1}, either favors fixing the other variables "off" or is indifferent about whether they are "on", "off" or "free" (Figure 3b). For an edge focused local learner, the set of possible focuses includes all the edges $\mathcal{L} \in \forall_{i<j \in N} E_{ij}$.

Effects

A commonly proposed heuristic for efficient search in the deterministic domains is to ask about the dimension that best divides the hypothesis space, eliminating the greatest possible number of options on average. This is variously known as "constraint-seeking" (Ruggeri & Lombrozo, 2014) or "the split half heuristic" (Nelson, Divjak, Gudmundsdottir, Martignon, & Meder, 2014). In the case of identifying the true deterministic ($w_S = 1$ and $w_B = 0$) causal model on N variables through interventions it turns out that the best split is achieved by querying the effects of a randomly chosen variable, essentially asking: "What does x do?" (Figure 3a)[9]. Formally we might think of this question as asking: which other variables (if any) are descendants of variable x in the true model? This a broader focus than querying the state of a single edge, but considerably simpler question than the global "which is the right causal model?" because the possibilities just include the different combinations of the other variables as effects (e.g. *neither*, *either* or *both* of y and z are descendants of x in a 3-variable model) rather than the superexponential number of model possibilities[10].

Relative to a chosen variable x, we can write an *effect focus* goal as maximizing the expectation of

$$\Delta H(\mathrm{De}(x)|\mathbf{d}, \mathbf{w}; \mathbf{c}) \tag{11}$$

where $\mathrm{De}(x)$ is is the set of x's direct or indirect descendants. This focus does not depend on b^{t-1}. This goal results in a preference for fixing the target node "on" (e.g. $\mathrm{Do}[x = 1]$) and leaving the rest of the variables free to vary (Figure 3b). For an effect focused local learner, the set of possible focuses includes all the nodes $\mathcal{L} \in \forall_{i \in X} \mathrm{De}(X\{i\})$.

Confirmation

Another form of local test, is to seek evidence that would confirm or refute the current hypothesis, against a single alternative "null" hypothesis. Confirmatory evidence gathering is a ubiquitous psychological phenomenon (Klayman & Ha, 1989; Nickerson, 1998). Although confirmation seeking is widely touted as a bias, it can also be shown to be optimal, for exam-

[9] This is also the most globally informative type of test relative to a uniform prior in all of the noise conditions we consider in the current paper

[10] The number of directed acyclic graphs on N nodes, $|\mathcal{M}|_N$, can be computed with the recurrence relation $|\mathcal{M}|_N = \sum_{k \in N} (-1)^{k-1} \binom{N}{2} 2^{k(N-k)} |\mathcal{M}|_{N-1}$ (see Robinson, 1977)

ple under deterministic or sparse hypotheses spaces or peaked priors (Austerweil & Griffiths, 2011; Navarro & Perfors, 2011).

Accordingly, Coenen et al. (2015) propose that causal learners adopt a "positive test strategy" when distinguishing causal models. They define this as a preference to "turn on" a parent component of one's hypothesis – observing whether the activity propagates to the other variables in the way that this hypothesis predicts. They find that people often intervene on suspected parent components, even when this is uninformative, and do so more often under time pressure. In Coenen et al's tasks, the goal was always to distinguish between two hypotheses, so their model assumed people would sum over the number of descendants each variable had under each hypotheses and turn on the component that had the most descendants on average. However, this does not generalize to the current, unrestricted, context where all variables have the same number of descendants if you average over the whole hypothesis space. However, Steyvers et al (2003) propose a related *rational test model* that selects interventions with a goal of distinguishing a single current hypothesis from a null hypothesis that there is no causal connection.

Following Steyvers et al. (2003), for a *confirmatory* focus we consider interventions expected to best reduce uncertainty between the learner's current hypothesis b^{t-1} and a null b^0 in which there are no connections (Figure 3a).

$$\Delta H(\{b^t, b^0\}|b^{t-1}, \mathbf{d}, \mathbf{w}; \mathbf{c}) \tag{12}$$

This goal results in a preference for fixing on the root node(s) of the target hypothesis (Figure 3c ii, noting the confirmation focus favours $\mathrm{Do}[x = 1, y = 1]$ here). The effectiveness of confirmatory focused testing depends on the level of noise and the prior, becoming increasingly useful later once the model being tested has sufficiently high prior probability. For a confirmation focused learner there is always just a single local focus.

Implications of the schema

The local uncertainty schema implies that intervention choice depends on two separable stages. Thus it accommodates the idea that a learner might be poor at choosing what to focus on but good at selecting an informative intervention relative to their chosen focus. It also allows that we might understand differences in learners' intervention choices as consequences of the types of local focus they are inclined or able to focus on. Learners cognizant of the limitations in their ability to incorporate new evidence might choose to focus their intervention on narrower questions (i.e. learning about a single edge at a time) while others might focus too broadly and fail to learn effectively. In the current work we will fit behavior assuming that learners choose between these local focuses, using their patterns to diagnose which local focuses they include in their option set $\mathcal{L}$, which of these they choose on a given test l^t and finally how these choices relate to their final performance.

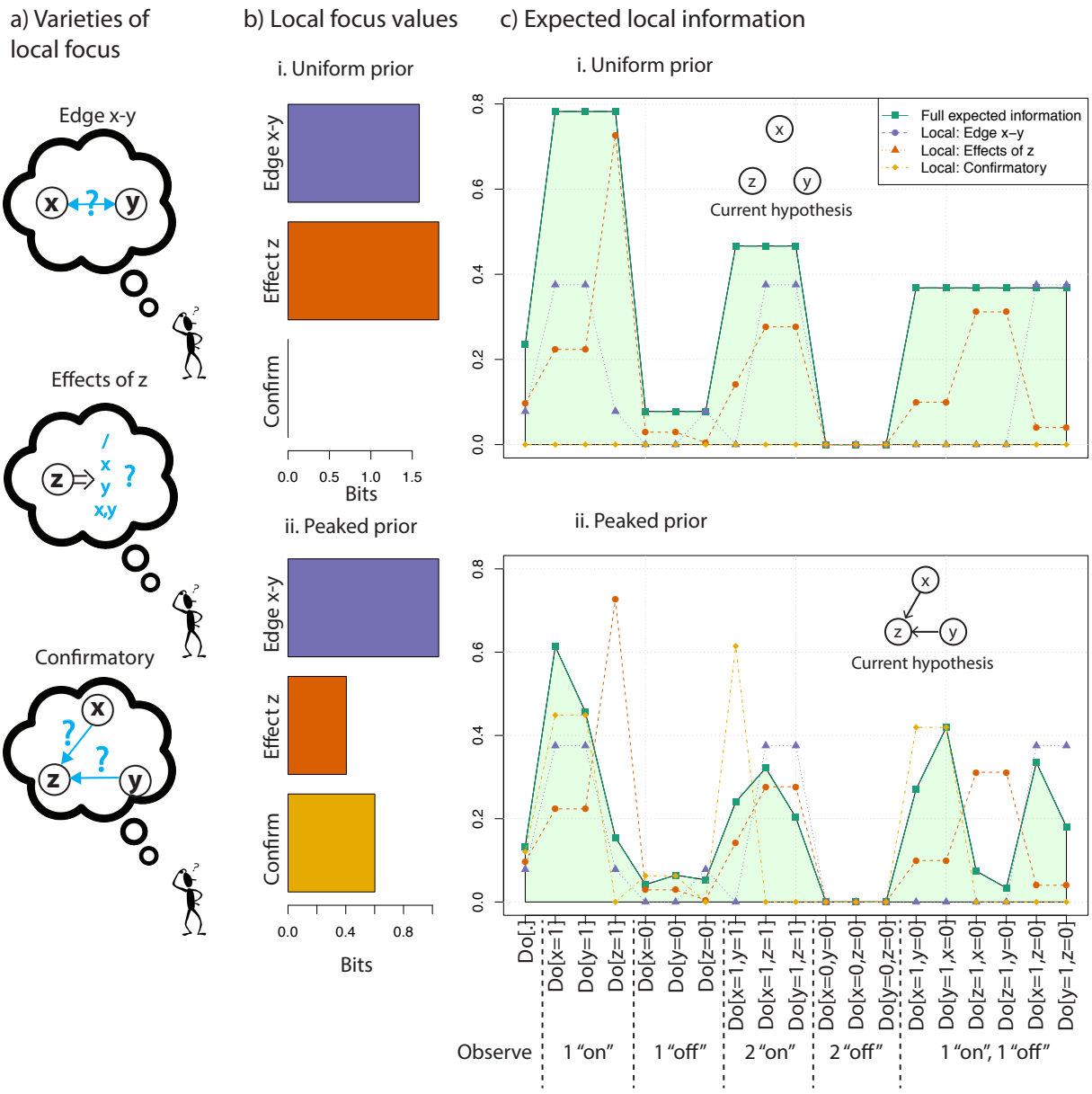

Figure 3: An illustrative example of local focused uncertainty minimization a) Three possible "local" focuses. b) The value of these choices of focus according to their current uncertainty Equation 6 (i.) at the start of learning and (ii.) after several tests have been performed. Note that uncertainty is measured with Shannon entropy based on the local possibilities and $\mathcal{D}_r^t$ and that confirmation is undefined at the start of learning where both current and null hypothesis are that there are no connections in the model. c) Expected value of 19 different interventions assuming: global expected information gain from the true prior (green squares, and shaded), effects of z focus (red circles), the relationship between x and y (blue triangles) and confirming b^{t-1} (yellow diamonds), assuming a uniform prior over the requisite possibilities and a known w_S and w_B of .85 and .15.

Comparing model predictions to experiments

The *Neurath's ship* framework we have introduced has two distinct signatures. Making only local edits from a single hypothesis results in sequential dependence. Making these edits by local resampling leads to aggregate behavior that can range between probability matching and hill climbing–which can give better short term gains but with a tendency to get stuck in local

optima. Two of the other heuristics also lead to sequential dependence. *Win-stay lose-sample* predicts all-or-none dependence whereby learners' judgments will either stay the same or jump to a new location that depends only on the posterior. *Simple endorsement* also predicts recency, although is distinguished by its failure to separate direct from indirect effects of interventions, leading to a different pattern of structural change.

In terms of interventions, if participants are locally focused, we expect their hypotheses to deviate from optimal predictions in ways that can be accommodated by our local uncertainty schema, i.e. selecting interventions that are more likely to be targeted toward local rather than global uncertainty. If learners do not maintain the full posterior, we expect their intervention distributions to be relatively insensitive to the evidence that has already been seen, while still being locally informative. If people disproportionately focus on identifying *effects*, we expect to see relatively unconstrained interventions with one variable fixed "on" at a time. If people focus on individual *edges* we expect more constraining interventions with more variables fixed "off". If *confirmatory* tests are employed, we expect to see more interventions on putative parents than on child nodes.

We first compare the predictions of our framework to existing data from Bramley, Lagnado, and Speekenbrink (2015). We then report on three new experiments designed to further test the specific predictions of our framework.

Bramley, Lagnado, and Speekenbrink (2015)

In Bramley, Lagnado, and Speekenbrink (2015), participants interacted with five probabilistic causal systems involving 3 variables (see Figure 4a), repeatedly selecting interventions (or tests) to perform in which any number of the variables are either fixed "on" or "off", while the remainder are left free to vary. The tests people chose, along with the parameters $\mathbf{w}$ of the true underlying causal model, jointly determined the data they saw. In this experiment w_S was always .8 and w_B was always .1. After each test, participants registered their best guess about the underlying structure. They were incentivised to report their best guess about the structure, through receipt of a bonus for each causal relation (or non-relation) correctly registered at the end. There were three conditions: *no information* (N=79) was run first. After discovering that a significant minority of participants performed at chance, condition *information* (N=30), added a button that participants could hover over and remind themselves of the key instructions during the task (the noise, strengths, the goal) and condition *information + summary* (N=30) additionally provided a visual summary of all previous tests and their outcomes.[11] Participants could draw cyclic causal models if they wanted (e.g. $x \rightarrow y \rightarrow z \rightarrow x$) and were not forced to select something for every edge from the start but instead could leave some or all of the edges as "?". Once a relationship was selected they could not return to "?". The task is available online at `http://www.ucl.ac.uk/lagnado-lab/el/ns_sup`.

[11] In the paper this was reported as two experiments, the second with two between-subjects conditions. They share identical structure they were subsequently analyzed together. Therefore we do the same here, reporting as a single experiment with three between-subjects conditions.

Comparing judgment patterns

We compared participants' performance in Bramley, Lagnado, and Speekenbrink (2015) to that of several simulated learners. *Posterior* draws a new sample from the posterior for each judgment. *Random* simply draws a random graph on each judgment. *Neurath's ship* follows the procedure detailed in the previous section, beginning with its previous judgment (b^{t-1}, or an unconnected model at t=1) and reconsidering one edge at a time based on the evidence gathered since its last change $\mathcal{D}_r^t$ for a small number of steps after observing each outcome. We illustrate this with a simulation with a short mean search length λ of 1.5 and behavior ω of 10 corresponding moderate hill climbing. *Win-stay, lose-sample* sticks with the previous judgment with probability $1 - P(D^t|b^{t-1}\mathbf{w}; C^t)$ or alternatively samples from the full posterior. The *simple endorser* always adds edges from any intervened-upon variables to any activated variables on each trial, and removes them from any intervened-upon variables to any non-activated variables, overwriting any edges going in the opposing direction. Participants' final accuracy in Bramley, Lagnado, and Speekenbrink (2015) was closest to the *Neurath's ship* as is clear in Figure 5a and b. That the Neurath's ship simulation unperformed participants in condition *information + summary* is to be expected since these participants were given a full record of past tests while *Neurath's ship* uses only the recent data.

Additionally, participants' online judgments exhibited sequential dependence. This can be seen in Figure 5b comparing the distribution of edits (bars) to the markedly larger shifts we would expect to see assuming random or Bayesian posterior sampling on these trials (black full and dotted lines). The overall pattern of edit distances from judgment to judgment is commensurate with those produced by the *Neurath's ship* procedure (red line), but also, here by *win-stay, lose-sample* (blue line) and *simple endorser* (green line) simulations.

Comparing intervention patterns

To compare intervention choices to global and locally driven intervention selection, we simulated the task with the same number of simulations as participants, stochastically generating the outcomes of the simulations' intervention choices according to the true model and true $\mathbf{w}$ (which the participants knew). Simulated efficient active learners would perfectly track the posterior and always select the greediest intervention (as in Equation 4).

We also compared participants' interventions to those of several other simulated learners, each restricted to one of the three types of local focus introduced in Section 4 ('edge'. 'effects' or 'confirmation').[12] When one of the simulated learners did not generate a unique best intervention, it would sample uniformly from the joint-best interventions according to that criterion. The results of the simulations are visualized in Figure 5c and d.

Participants' intervention choices in Bramley, Lagnado, and Speekenbrink (2015) were clearly more informative than random selection but less so than ideal active learning. This is evident in Figure 5c comparing participants (bars) to simulations of ideal active learning (black circles) and random intervening (black squares), and in Figure 5d comparing the participants

[12]We assumed these tests were chosen based on a uniform prior over the options considered. We used the latest most probable judgment $\text{argmax}p(M|D^{t-1}, \mathbf{w})$ in place of a current hypothesis b^{t-1} for edge focused and confirmatory testing so as not to presuppose a particular belief update rule in assessing intervention selection.

(red lines) to the ideal active learning (pink lines) and random intervening (blue lines) simulations. Furthermore, the informativeness of participants' interventions is in the range of the simulations of any of the three local foci (yellow, green and blue lines).

As we see in Figure 5d, idealized active learning favored fixing one variable on at a time ($\mathrm{Do}[x = 1]$, $\mathrm{Do}[y = 1]$ etc, hereafter called "one-on" interventions) for the majority of tests. It always chose "one-on" for the first few tests but would sometimes select controlled (e.g. $\mathrm{Do}[x=1, y=0]$) tests on later tests when the remaining uncertainty was predominantly between direct and indirect causal pathways as in between chain, fork and fully connected structures.

Locally driven testing had different signatures depending on the focus. The *edge* focused simulation would fix the component at one end of their edge of interest "on" and leave the component at the other end "free". What it did with the third component depended on its latest judgment about the network. If, according to b^{t-1}, another component was a cause of the component that was left free-to-vary, the simulation favored fixing it "off". Otherwise, it did not distinguish between "on", "off" or "free" choosing one of these at random. The resulting pattern is a spread across "one-on", "two-on" and "one-on, one-off" tests with a bias toward controlled "one-on, one-off" tests. The effects focused learner always favored "one-on" interventions. The *confirmation* focused tester would generally fix components with children in b^{t-1} on, and leave components with parents in b^{t-1} free. This led to the choice of a mixture of "one-on" and "two-on" interventions.

Like the *ideal* or the *effects* focused simulations, participants in Bramley, Lagnado, and Speekenbrink (2015) strongly favored "one-on" tests. Consistent with *confirmatory* testing, components with at least one child according to the latest hypothesis b^{t-1} were more likely to be fixed "on" than components believed to have no children (60% compared to 56% of the time $t(24568) = 3.2, p = .001$).[13] Participants' intervention selections were markedly less dynamic across trials than those of the efficient learner. For example, the proportion of single (e.g. $[x=1]$) interventions decreased only fractionally on later tests, dropping from 78% to 73% from the first to the last test.

[13] We ran the name number of simulated learners as participants in each experiment and condition to facilitate statistical comparison.

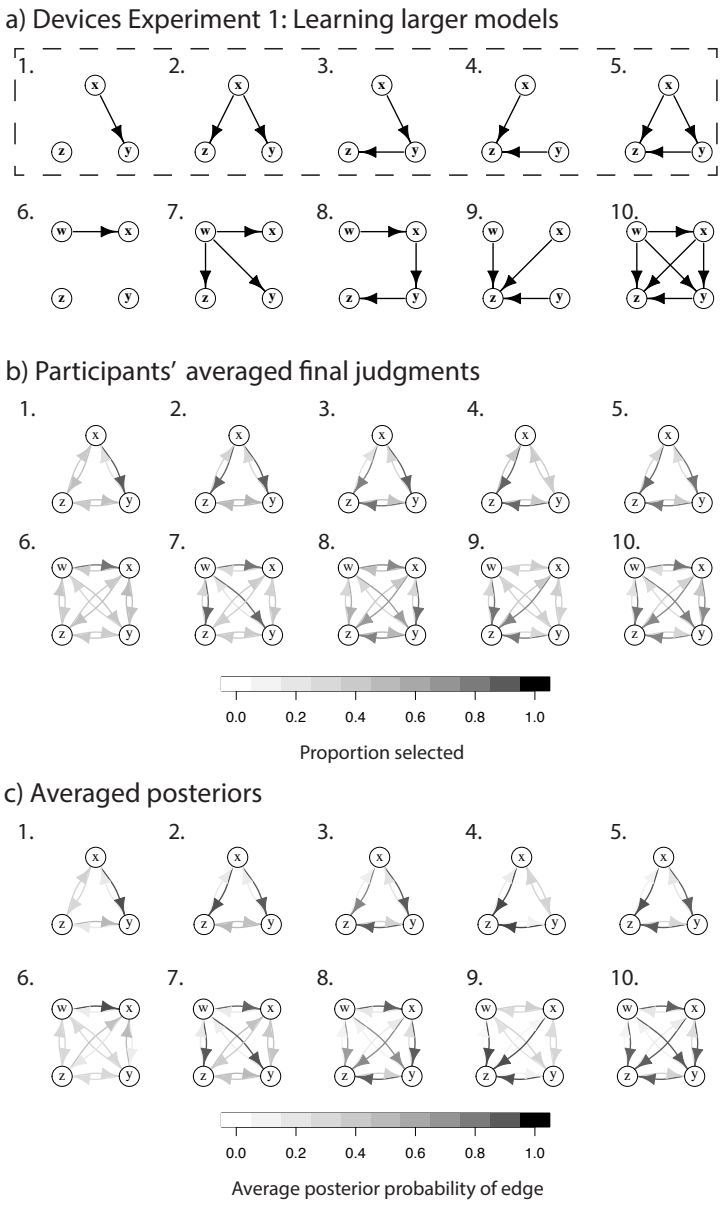

Figure 4: The true models from Experiment 1: Learning larger models, and visualisation of averaged judgments and posteriors. a) The problems faced by participants. Dashed box indicates those that also appeared in Bramley, Lagnado, and Speekenbrink (2015). b) Averaged final judgments by participants. Darker arrows indicate that a larger proportion of participants marked this edge in their final model. c) Bayes-optimal final marginal probability of each edge in $P(M|D^T, E^T, \mathbf{w})$, averaged over participants' data.

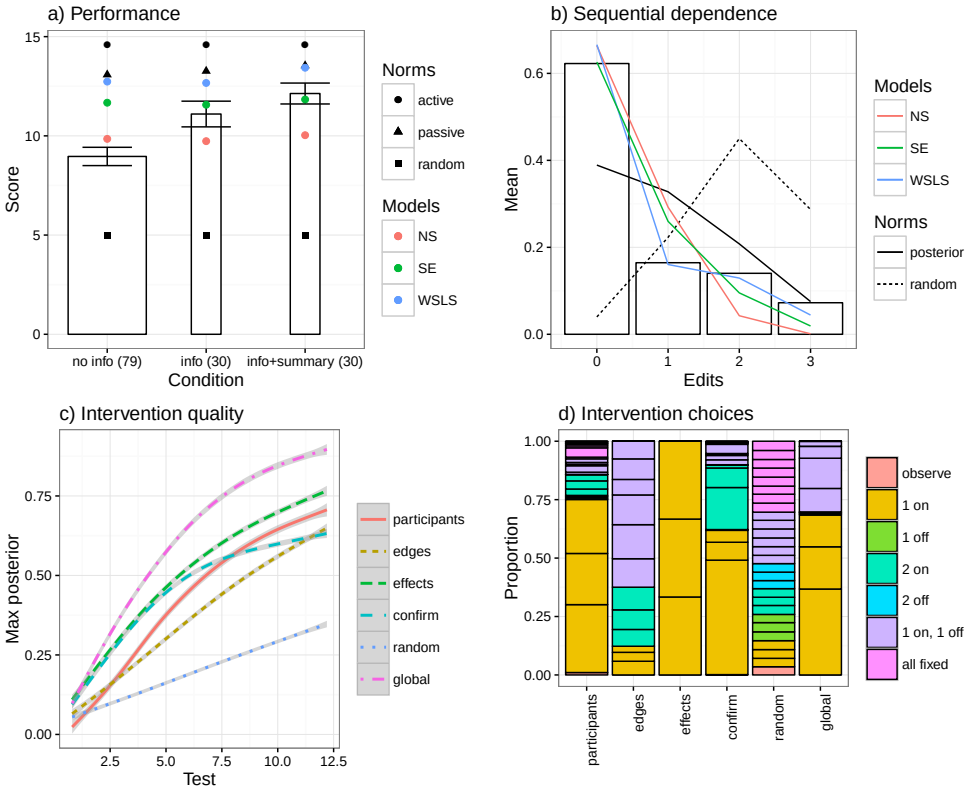

Figure 5: Bramley, Lagnado, and Speekenbrink (2015); performance and interventions. a) Accuracy by condition. Bars show participant accuracy by condition, and points compare with the models, bar widths visualize the number of participants per condition. b) Sequential dependence. The number of edits made by participants between successive judgments, bars give proportion of participants updates with different numbers of edits, lines compare with the models. c) Quality of participants' and simulated learners' intervention choices measured by the probability that an ideal learner would guess the correct model given the information generated. The plot shows values smoothed with R's `gam` function and the gray regions give 99% confidence intervals. The proportion of interventions of different types chosen by participants as compared to simulated learners. observe = $\mathrm{Do}[\varnothing]$, 1 on = e.g. $\mathrm{Do}[x=1]$, 1 off = e.g. $\mathrm{Do}[x=0]$ and so on. All fixed = e.g. $\mathrm{Do}[x=0, y=1, z=0]$.

Motivating the new experiments

In analyzing Bramley, Lagnado, and Speekenbrink (2015), we found patterns of judgments and interventions broadly consistent with our framework. However, the conclusions we can draw from this data alone are somewhat limited. Firstly, the problems participants faced did not strongly delineate our *Neurath's ship* proposal from other proposed approximations, namely the approximate *win-stay, lose-sample* or the heuristic *simple endorsement* which also predicted similar patterns of accuracy and sequential dependence.

Similarly, in terms of interventions, participants' strong preference for "one-on" interventions was consistent with local effect-focused testing. However, "one-on" interventions were also the globally most informative choices for the majority of participants' trials, especially early during learning. Thus, we cannot be confident what participants focused on when selecting their interventions.

Methodologically also, several aspects of Bramley, Lagnado, and Speekenbrink (2015) are suboptimal for testing our framework. Participants were allowed to leave edges unspecified

until the the last test and could also draw cyclic models, both of which complicated our analyses. Furthermore, participants had 12 tests on each problem, allowing an idealized learner to approach certainty given the high w_S and low w_B, and for a significant minority of people to perform at ceiling. These choices limit the incentive for participants to be efficient with their interventions. Additionally, participants were only incentivised to be accurate with their final judgment, meaning we cannot be confident that intermediate judgments always represented their best and latest guess about the model. Finally, participants were not forced to update all their edges after each test, meaning that lazy responding could be confused with genuine sequential dependence of beliefs.

Next, we report on two new experiments that build on the paradigm from Bramley, Lagnado, and Speekenbrink (2015), making methodological improvements, while also exploring harder more revealing problems, and eliciting additional measures, all with the goal of better distinguishing our framework from competitors.

Experiment 1 explores learning in more complex problems than in Bramley, Lagnado, and Speekenbrink (2015), with more variables and a range of strengths w_S and levels of background noise w_B, and fewer interventions per problem. The increased complexity and noise provides more space and stronger motivation for the use of approximations and heuristics. Furthermore, the broader range of possible structures and intervention choices increases the discriminability of our framework from alternatives such as *win-stay, lose-sample* and *simple endorsement*, while the shorter problems avoid ceiling effects and ensure participants choose interventions carefully. To ensure participants register their best and latest belief at every time point, we also incentivize participants through their accuracy at random time points during learning. To eliminate the possibility of lazy responding biasing results in favor of *Neurath's ship*, we force participants to mark all edges anew after every test without a record of their previous judgment as a guide.

Experiment 2 inherits the methodological improvements, compares two elicitation procedures, and also takes several additional steps. In the previous studies, participants were pretrained on strength w_S and background noise w_B. This will not generally be true; learners will normally have to take into account their uncertainty about these sources of noise during inference. Therefore, Experiment 2 focuses on cases where participants are not pretrained on $\mathbf{w}$. Additionally, our framework makes predictions about participants' problem representation that go beyond how it should manifest in final structure judgments and intervention choices. Specifically, our local intervention schema proposes that people focus on subparts of the overall problem during learning, switching between these by comparing their current local uncertainty. Experiment 2 probes these assumptions by asking learners for confidence judgments about the edges in the model during learning, and eliciting free explanations of what interventions are supposed to be testing. When we go on to fit our framework to individuals in the final section of the paper, we are able to code up these free responses in terms of the hypotheses they refer to and compare them to the focuses predicted by our local uncertainty schema.

Experiment 1: Learning larger causal models

Our first new experiment looks at learning in harder problems with a range of w_S and w_B and a mixture of 3- and 4-variable problems, asking whether we now see a clearer signature of *Neurath's ship*, *simple endorsement* or *win-stay,lose-sample* style local updating or of local focus during interventions selection.

Methods

Participants

120 participants (68 male, mean±SD age 33 ± 9) were recruited from Amazon Mechanical Turk[14], split randomly so that 30 performed in each of 4 conditions. They were paid \$1.50 and received a bonus of 10c per correctly identified connection on a randomly chosen test for each problem ($\max = \$6.00$, mean±SD $\$3.7 \pm 0.65$). The task took an average of 44 ± 40 minutes.

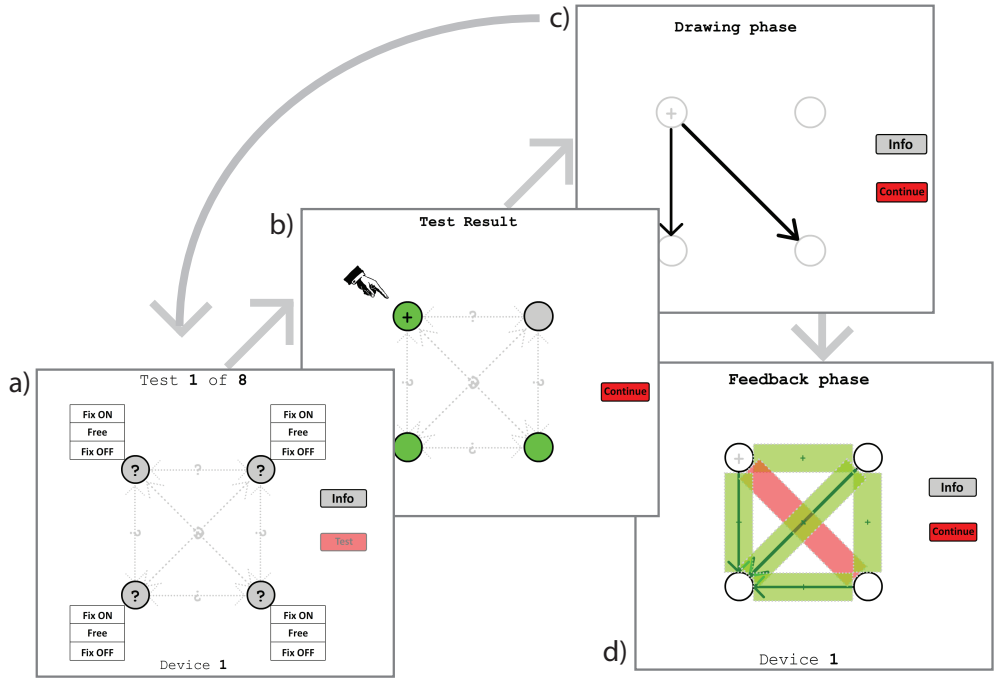

Figure 6: Experimental procedure. a) Selecting a test b) Observing the outcome c) Updating beliefs d) Getting feedback.

Design

This study included the five 3-variable problems in Bramley, Lagnado, and Speekenbrink (2015) plus five additional 4-variable problems (see Figure 4a). There were problems exemplifying three key types of causal structure: forks (diverging connections), chains (sequential connections) and colliders (converging connections). Within these, the sparseness of the causal connections varied between a single connection (devices 1 and 6) and fully connected (devices

[14]Mechanical Turk (`http://www.mturk.com/`) is a web based platform for crowd-sourcing short tasks widely used in psychology research. It offers an well validated (Buhrmester, Kwang, & Gosling, 2011; Crump, McDonnell, & Gureckis, 2013; Gosling, Vazire, Srivastava, & John, 2004; Hauser & Schwarz, 2015; Mason & Suri, 2012) subject pool, diverse in age and background, suitable for high-level cognition tasks.

5 and 10).

There were two different levels of causal strength $w_S \in [.9, 0.75]$ and two different levels of background noise $w_B \in [.1, .25]$ making $2 \times 2 = 4$ between-subjects conditions. For instance, in condition $w_S = .9; w_B = .1$ the causal systems were relatively reliable, with nodes rarely activating without being intervened on, or caused by, an active parent, and connections rarely failing to cause their effects. Meanwhile, in condition $w_S = 0.75; w_B = 0.25$ the outcomes were substantially noisier, with probability .25 that a variable with no active parent would activate, compared to a probability $1 - (1 - .75)(1 - .25) = 0.81$ for a variable with one active parent.

Procedure

The task interface was similar to that in Bramley, Lagnado, and Speekenbrink (2015). Each device was represented as several gray circles on a white background (see Figure 6). Participants were told that the circles were components of a causal system of binary variables, but were not given any further cover story. Initially, all components were inactive and no connection was marked between them. Participants performed tests by clicking on the components, setting them at one of three states "fixed on", "fixed off" and "free-to-vary", then clicking "test" and observing what happened to the "free-to-vary"components as a result. The observations were of temporary activity (graphically, activated components would turn green and wobble).

As in Bramley, Lagnado, and Speekenbrink (2015), participants registered their best guess about the underlying structure after each test. They did this by clicking between the components to select either no connection, or a forward or backward connection (represented as black arrows). Participants were incentivised to be accurate, but unlike in Bramley, Lagnado, and Speekenbrink (2015), payments were based on randomly selected time points rather than the final judgments.

Participants completed instructions familiarizing them with the task interface; the interpretation of arrows as (probabilistic) causal connections; the incentives for judgment accuracy. To train $\mathbf{w}$, participants were told explicitly and then shown unconnected components and forced to test them several times. The frequency with which the components activated reflected the true background noise level. They were then shown a set of two-component causal systems in which component "x" was a cause of "y", and were forced to test these systems several times with component x fixed on. This indicated that the frequency with which y activated reflected the level of w_S combined with the background noise they had already learned.

After completing the instructions, participants had to answer four comprehension check questions. If they got any wrong they had to go back to the start of the instructions and try again. Then, participants solved a practice problem randomly drawn from the problem set. They then faced the test problems in random order, with randomly oriented unlabeled components. They performed six tests on each three variable problem, and eight tests on each four variable problem. After the final test for each problem they received feedback telling them the true connections.

To ensure that participants' judgments were always genuine directed acyclic graphs, participants were told in the instructions that the true causal structure would not contain a loop. Unlike in Bramley, Lagnado, and Speekenbrink (2015), if participants tried to draw a model

containing a cyclic structure they would see a message saying "you have drawn connections that make a loop, change or remove one to continue".

As in Bramley, Lagnado, and Speekenbrink (2015) conditions *information* and *information + summary*, participants could hover their mouse over a button for a reminder of the key instructions during the task, but unlike condition *information + summary*, they saw no record of their previous tests and outcomes.

The task can be tried out at `http://www.ucl.ac.uk/lagnado-lab/el/ns_sup`.

Results and discussion

Judgments

In spite of the considerably greater noise and complexity than Bramley, Lagnado, and Speekenbrink (2015), participants performed significantly above chance in all four conditions (comparing to chance performance of $\frac{1}{3}$, participants scores differed significantly by t-test with $p < .001$ for all four conditions). They also significantly underperformed a Bayes optimal observer ($p < .001$ for all four conditions, Figure 7a). Performance declined as background noise w_B increased $F(1, 118) = 4.3, \eta^2 = .04, p = .04$ but there was no evidence for a relationship with strength w_S $F(1, 118) = 2.7, \eta^2 = .04, p = 0.1$. Judgment accuracy was no lower for four compared to three variable problems $t(238) = 0.76, p = 0.44$. Table 1 shows accuracy by device type across all experiments. Accuracy differed by device type $\chi^2 = (4) = 22, p < .001$. Consistent with the idea that people struggle most to distinguish the chain from the fork or the fully connected model, accuracy was lowest for chains (devices 3; 8) and second lowest for fully connected (5; 10) models.

In all four conditions, participants' final accuracy was closer to that of the *Neurath's ship* simulations than the *simple endorser*, *win-stay, lose sample* or *random* responder or ideal (*passive*) responding (Figure 7a).[15]

Table 1: Proportion of Edges Correctly Identified by Device Type in All Experiments

Experiment	Variables	Single	Fork	Chain	Collider	Fully-connected	No connection
Bramley et al (2015)	3	0.69	0.63	0.66	0.71	0.67	
Exp 1: Learning larger models	3	0.57	0.56	0.51	0.6	0.55	
Exp 1: Learning larger models	4	0.56	0.58	0.48	0.57	0.49	
Exp 2: Unknown strengths	3	0.62	0.65	0.61	0.6	0.61	0.61
All		0.61	0.61	0.57	0.62	0.58	0.61

Sequential dependence

Table 2 summarizes the number of edits (additions, removals or reversals of edges) participants made between each judgment in all experiments. Inspecting the table and Figure 7b we see participants judgments (both high and low performing) show a pattern of rapidly decreasing probability for larger edit distances mimicked by both *Neurath's ship* and *simple endorsement*

[15] On the rare occasions where the *simple endorser* procedure would induce a cycle (0.4% of trials), the edges were left in their original state.

Table 2: Edit Distance Between Consecutive Judgments in All Experiments.

Experiment	Var	Participants		Random		SE		WSLS		NS		Posterior	
		M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Bramley et al (2015)	3	0.66	0.97	1.99	0.83	0.51	0.74	0.55	0.88	0.38	0.57	1.02	0.93
Exp 1: Learning larger models	3	0.92	1.01	1.99	0.83	0.44	0.64	0.73	0.98	0.50	0.63	1.47	0.94
Exp 1: Learning larger models	4	1.69	1.63	3.96	1.16	0.65	0.88	1.64	1.78	0.55	0.68	2.94	1.40
Exp 2: Unknown strengths (remain)	3	0.73	0.85	2.02	0.81	0.43	0.62	1.06	1.13	0.59	0.69	1.86	0.85
Exp 2: Unknown strengths (disappear)	3	1.02	0.99	2.02	0.81	0.43	0.62	1.06	1.13	0.59	0.69	1.86	0.85
All		0.99	1.09	2.40	0.89	0.50	0.71	0.99	1.17	0.51	0.65	1.80	1.00

NOTE: Var = number of variables. NS = *Neurath's ship* simulations with $\lambda = 1.5$ and $\omega = 10$, WSLS = *Win-stay, lose-sample* simulations, SW = *Simple Endorser* simulations, M=mean, SD = standard deviation.

simulations. In contrast, *random* or *posterior* sampling lead to quite different signatures with larger jumps being more probable. Choices simulated from *Neurath's ship* and *simple endorsement* were more sequentially dependent than participants' on average but have the expected decreasing shape. *Win-stay, lose sample* produces a different pattern with a maximum at zero changes but a second peak in the same location as for *posterior sampling* but has an average edit distance very close to that averaged over participants. However, we expect any random or inattentive responding to inflate average edit distances, and indeed find a strong negative correlation between edit distance and score $F(1, 118) = 34, \beta = -6.7, \eta^2 = .34, p < .001$. A simple way to illustrate this is to compare the edits of higher and lower performers. Scores of $\frac{22}{45}$ or more differ significantly from chance performance (around $\frac{15}{45}$) by χ^2 test. The 79 participants that scored 22 or more made markedly smaller edits than those that scored under 22 (0.85 ± 0.95 compared to 1.3 ± 1.12 for three variable, and 1.4 ± 1.5 compared to 2.4 ± 1.8 for four variable problems), putting the clearly successful participants patterns closer to the "Neurath's ship" and "simple endorser" simulations. Additionally, we expect individual differences in search length λ under the *Neurath's ship* model and here only simulate assuming a mean search length of 1. Aggregating over a wider set of simulated learners with different capacities to search for updates would lead to a heavier-tailed distribution of edit distances that would resemble the participants' choices more faithfully.

Interventions

Globally focused active learning favored a mixture of "one-on" and "one-on, one-off" interventions (and several others including "one-on, two-off" in the four variable problems). The number and nature of the fixed variables it favored depended strongly on the condition, favoring fixing more variables off when w_S was high. It would also shift dramatically over trials always favoring "one-on" interventions for the first trials but these dropping below 50% of choices by the final test. Participants' choices were much less reactive to condition or trial. There were no clear differences in intervention choices by condition (see supplementary figures available at `http://www.ucl.ac.uk/lagnado-lab/el/ns_sup`) but participants were a little more likely to select "one on" interventions on their first test 57% compared to their last 50% test $t(238) = 1.7, p = .01$. Like the *ideal* or the *effects* focused simulations, and like in Bramley, Lagnado, and Speekenbrink (2015), learners favored "one-on" tests. However, in line with an *edge* or *confirmation* they also selected a substantial number of "two-on" and "one-on, one-

off" interventions, doing so on early as well as late tests while the ideal learner only predicted using "one-on, one-off" tests on the last few trials. As in Bramley, Lagnado, and Speekenbrink (2015) and consistent with *confirmatory* testing, participants were more likely to fix "on" components with at least one child according to their latest hypothesis b^{t-1}: 49% compared to 30% $t(238) = 5.5, p < .001$. The overall pattern was not clearly consistent with any one local focus but might be consistent with a mixture of all three.

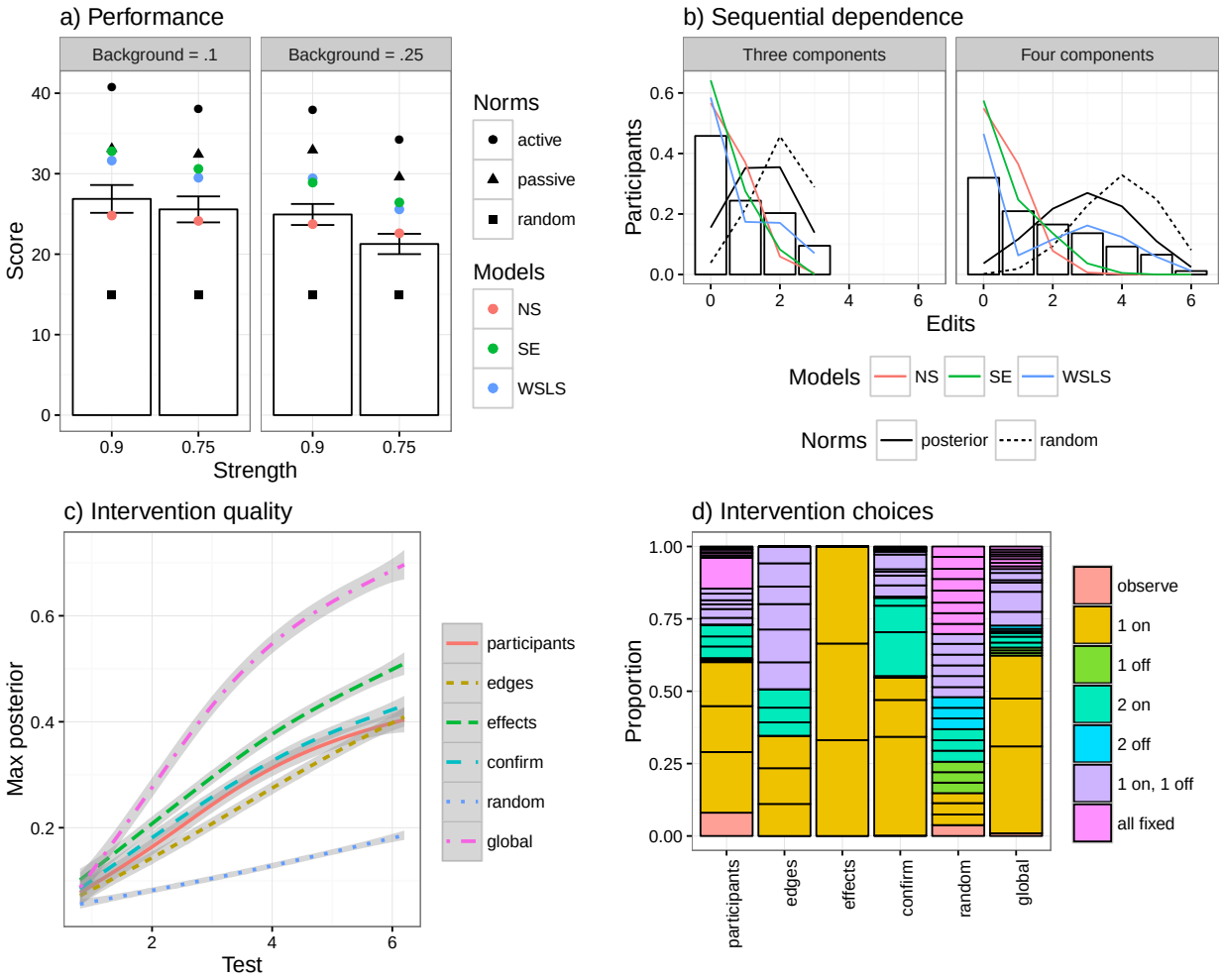

Figure 7: Experiment 1: Learning larger causal models; performance and interventions. a) Accuracy by condition. Bars are participants and points compare with the models. b) Sequential dependence. Bars show the number of edits made by participants between successive judgments. Lines compare with the models. c) Quality of participants' and simulated learners' intervention choices in the three variable problems as in Figure 5b. d) The proportion of interventions of different types chosen by participants compared to simulated learners in the three variable problems, otherwise as in Figure 5d.

Experiment 2: Unknown strengths

In this experiment, we focused on cases where participants are not pretrained on $\mathbf{w}$ (see Appendix A for the computational level details of how to incorporate uncertainty over $\mathbf{w}$ in

model inference and intervention choice).[16]

We took advantage of the fact that participants would experience substantially greater uncertainty given ignorance about $\mathbf{w}$ to assess their ability to estimate local uncertainty based on recently observed data $\mathcal{D}_r^t$ in order to choose where to focus subsequent tests. This is central to any scheme for intervention selection. Thus, in Experiment 3, we elicited the participants' confidence about the edges in each judgment. If participants track local uncertainties based on recent evidence, we should expect these to correlate with uncertainties given $\mathcal{D}_r^t$. In particular, given the representation associated with *Neurath's ship*, we might also expect the local confidences to be evaluated while leaning on the rest of the model for support. This means they should reflect conditional uncertainty in the edge $H(E_{ij}|E_{\setminus ij}, \mathcal{D}_r^t; \mathcal{C}_r^t)$ more closely than the marginal uncertainty $H(E_{ij}|\mathcal{D}_r^t; \mathcal{C}_r^t)$ which involves averaging across all the possible states of the other edges.

We also elicited predictions about the outcome of each chosen test before the outcome was revealed. If participants maintained only a single hypothesis, we expected this to be reflected in their predictions. Thus, for a *Neurath's ship* learner, it would be predominantly the predictive distribution under their current hypothesis rather than the average across models.

Finally, in Bramley, Lagnado, and Speekenbrink (2015) and Experiment 1, participants' intervention selections showed hints of being motivated by a mixture of local aspects of the overall uncertainty, with overall patterns most consistent with focus on a mixture of different local aspects of uncertainty. To test this idea more thoroughly, in the final problem in Experiment 2 we explicitly probed participants' beliefs about their intervention choices through eliciting free responses which we go on to code and compare to our model predictions.

Methods

Participants

111 UCL undergraduates (mean $\pm$ SD age 18.7 ± 0.9, 22 male) took part in Experiment 2 as part of a course. They were incentivised to be accurate based on randomly selected trials as before, but this time with the opportunity to win *Amazon*™ vouchers rather than money. Participants were split randomly into 8 groups of mean size 13.8 ± 3.4, each of which was presented with a different condition in terms of the value of $\mathbf{w}$ and the way that they had to register their responses.

Design and procedure

Experiment 2 used the same task interface as the other experiments, but focused just on the three variable problems (devices 1-5) and an additional device (6) in which none of the components was connected (Figure 9). Like in Experiment 1, there were two causal strength conditions $w_S \in [0.9, 0.75]$ and two background noise conditions $w_B \in [0.1, 0.25]$. However, unlike Experiment 1, participants were not trained on these parameters, but only told that: *"the connections do not always work"*, and *"sometimes components can activate by chance"*.

To assess the extent the different reporting conditions drove lower sequential dependence

[16] Experiments 2 is also reported briefly in Bramley, Dayan, and Lagnado (2015) but without discussion of the intervention choices or current model.

in Bramley, Lagnado, and Speekenbrink (2015) relative to Experiment 1, we examined two reporting conditions between subjects: *remain* and *disappear*. In the *remain* condition, judgments stayed on the screen into the next test, so participants did not have to change anything if they wanted to register the same judgment at t as at $t-1$. In the *disappear* condition, the previous judgment disappeared as soon as participants entered a new test. They then had explicitly to make a choice for every connection after each test.

In addition to the structure judgments and interventions, we also elicited additional probability measures from participants. First, after selecting a test, but before seeing the outcome, participants were asked to predict what would happen to the variables they had left free. To do this they would set a slider for each variable they had left free to vary. The left pole of the slider was labeled "Sure off", the right pole "Sure on"and the middle setting indicated maximal uncertainty (Figure 8a). Second, after drawing their best guess about the causal model by setting each edge between the variables, participants were asked how sure they were about each edge. Again they would respond by setting a slider, this time between "Guess" on the left indicating maximal uncertainty and "Sure" on the right indicating high confidence that that edge judgment was correct (Figure 8b). Participants were trained and tested on interpretation of the slider extremes and midpoints in an additional interactive page during the instructions.

Participants faced the six devices in random order, with six tests per device followed by feedback as in Experiments 1 and 2. Then they faced one additional test problem. On this problem, the true structure was always a chain (Figure 9, device 7). On this final problem, participants did not have to set sliders. Instead, after they selected each test, but before seeing its outcome, they were asked why they had selected that intervention. Labels would appear on the nodes and participants were invited to *"Explain why you chose this combination of fixed and unfixed components. Use labels 'A' 'B', 'C' to talk about particular components or connections"* in a text box that would appear below the device. Responses were constrained to be at least 5 characters long. The chain (device 3) was chosen for this problem because in Bramley, Lagnado, and Speekenbrink (2015) and Experiments 1 and 2, participants often did not select the crucial $\mathrm{Do}[x=1, y=0]$ intervention that would allow them to distinguish a chain from a fully connected model (device 5) making this an interesting case for exploring divergence between participants' behavior and ideal active learning.

Finally, at the end of the experiment participants were asked to estimate the reliability w_S of the true connections: *"In your opinion, how reliable were the devices? i.e. How frequently would fixing a cause component ON make the effect component turn ON too?"* and the level of background noise w_B: *"In your opinion, how frequently did components activate by themselves (when they were not fixed by you, or caused by any of the device's other components)?"* by setting sliders between *"0% (never)"* and *"100% (always)"*.

A demo of Experiment 2 can be viewed at `http://www.ucl.ac.uk/lagnado-lab/el/ns_sup`.

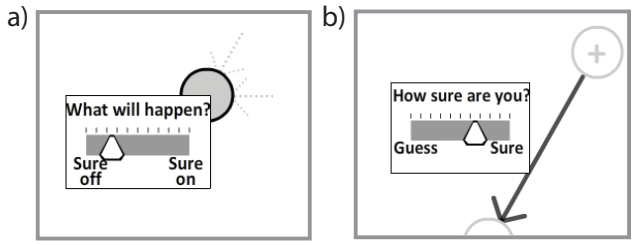

Figure 8: Exp 2: Unknown strengths; additional measures - a) Outcome expectation sliders b) Edge confidence sliders.

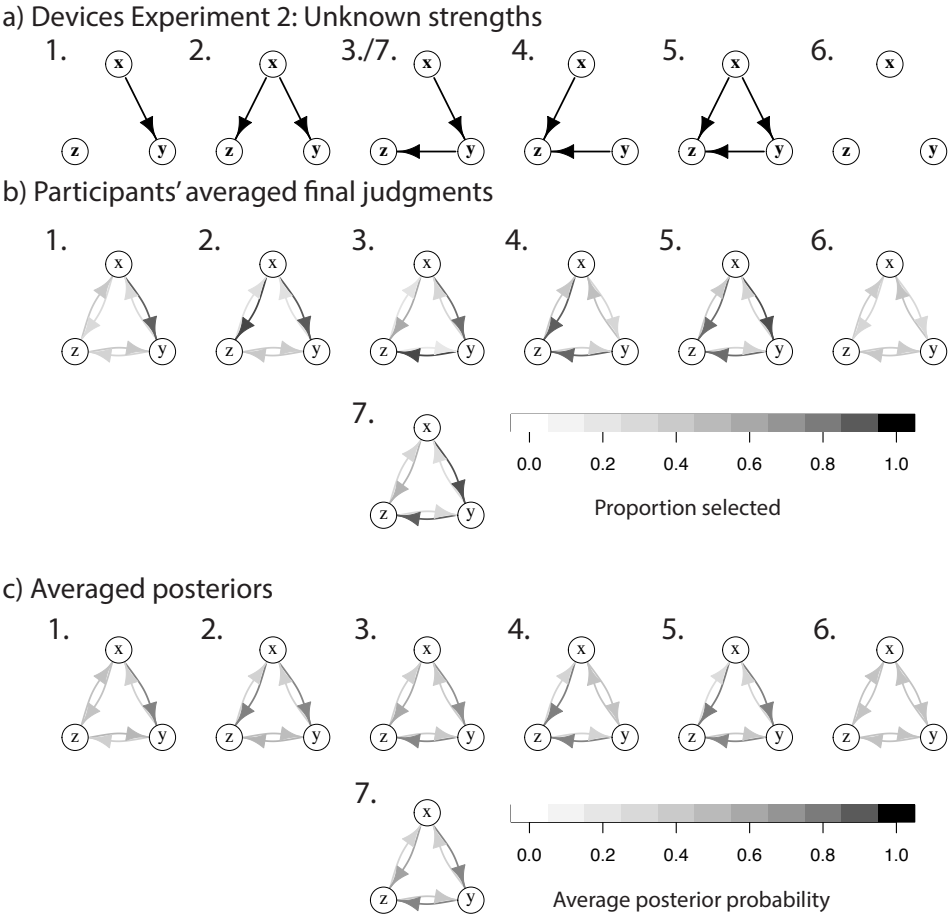

Figure 9: Experiment 2: Unknown strengths; true models and final judgments. a) The true models faced by participants. b) Weighted average final judgments by participants. Darker arrows indicate that a larger proportion of participants marked this link in their final model. Note that problem 7 was the a repeat of the chain (problem 3) with the write aloud protocol. c) Bayes-optimal marginal probability of each edge in $\int_{\mathbf{w}} P(M|D^T; C^T)p(\mathbf{w})\,\mathrm{d}\mathbf{w}$ averaged over participants' data assuming a uniform independent prior over $\mathbf{w} \in [0, 1]^2$.

Results and discussion

Judgments

As in the experiments where participants were trained on $\mathbf{w}$, accuracy was significantly higher than chance in all conditions (all 8 t statistics > 6.1 all p values < 0.001) and under-

performed a Bayes optimal observer observing the same data as them. Because the noise was unspecified, we explored several reasonable priors on $\mathbf{w}$ (always assuming that w_S and w_B were independent) when computing posteriors. Firstly, we considered a *uniform-uniform* prior that made no assumptions about either w_S or w_B (UU) where $\mathbf{w} \sim \text{Uniform}(0,1)^2$. We also considered a *strong-uniform* (SU) variant, following (Yeung & Griffiths, 2011), expecting causes to be reliable – $w_S \sim \text{Beta}(2,10)$, but making no assumptions about background noise – $w_B \sim \text{Uniform}(0,1)$. Additionally, we considered a *sparse-strong* (SS) variant following Lu et al (2008), encoding an expectation of high edge reliability – $w_S \sim \text{Beta}(2,10)$, and relatively little background noise – $w_B \sim \text{Beta}(10,2)$. The choice of parameter prior made little difference to the Bayes optimal observer's judgment accuracy. Thus participants significantly underperformed the Bayes optimal observer in all conditions regardless of the assumed prior, except for condition $w_S = 0.75; w_B = 0.1$, *remain*) under the SU prior, and $w_S = 0.75; w_B = 0.25$, *remain* under all three considered priors.

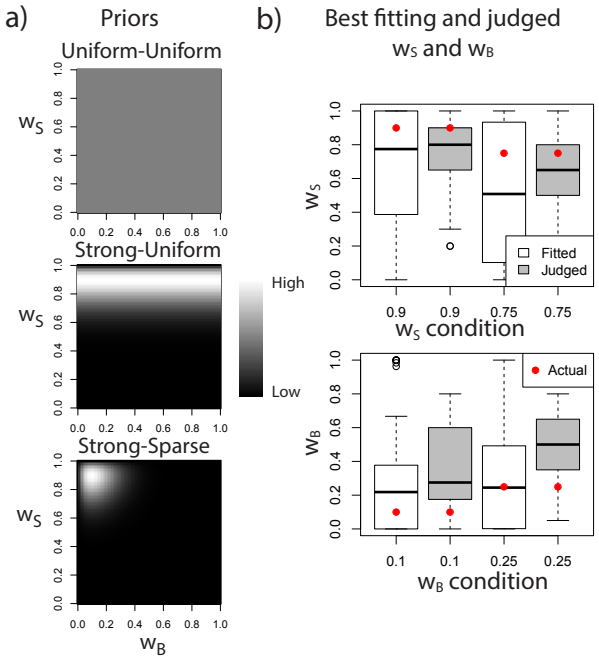

Figure 10: Experiment 2: Unknown strengths; priors and strength estimates. a) Visualizations of Uniform-Uniform, Strong-Uniform and Strong-Sparse priors on w_S and w_B b) Participants final judgments of amount of background noise (w_B) and strength (w_S), rescaled from 100 point scale to 0-1, and best-fitting w_S^* and w_B^* estimates assuming ideal Bayesian updating. Boxplots show the medians and quartiles.

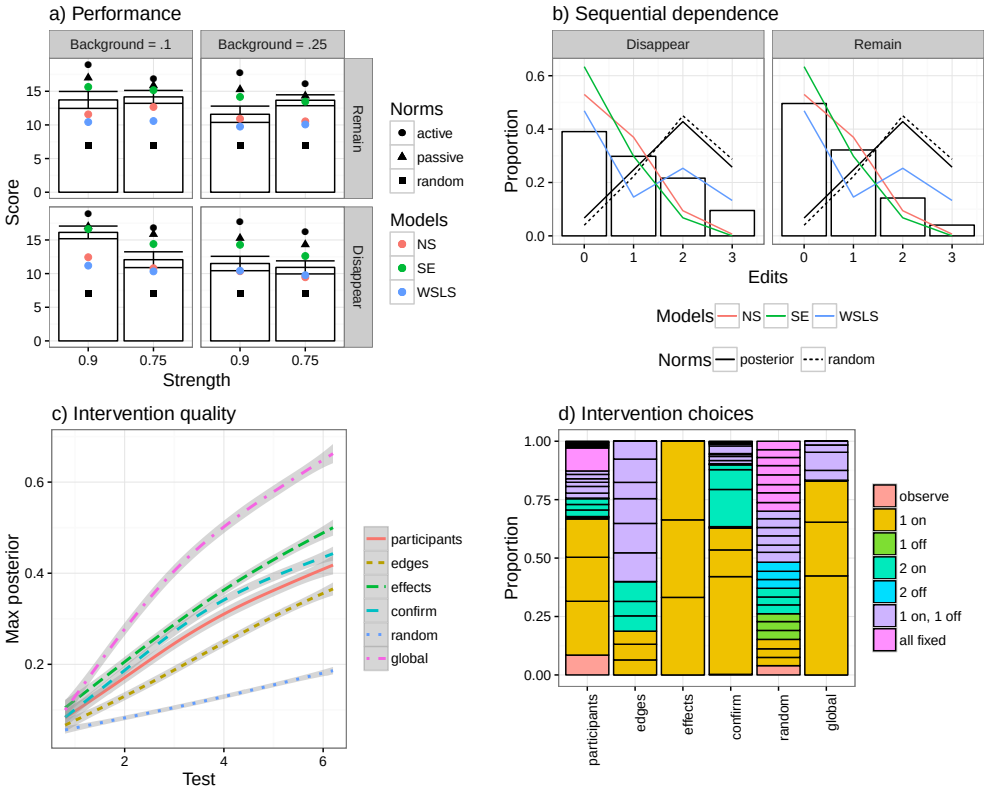

Figure 11: Experiment 2: Unknown strengths; performance and interventions. Subplots as in Figure 7.

Comparison with known strength experiments

Performance in Experiment 2 was comparable to the 3-variable problems in Experiment 1 where the underlying **w** conditions were identical. Mean accuracy was actually slightly higher 0.61 ± 0.21 compared to 0.56 ± 0.21 for the matched problems in Experiment 1, although not significantly so $t(229) = 1.9, p = .054$. This suggests that participants were able make reasonable structure judgments without knowledge of the exact parameters. We found that participants' final judgments of w_S and w_B and best fitting estimates assuming rational updating w_S^* and w_B^* suffered bias and variance (Figure 10 b).[17]

As with Experiments 1 and 2, participants were not affected by the reliability of the connections themselves w_S $t(106) = 0.88, p = 0.37$ but were affected by higher levels of background noise w_B $t(108) = 2.7, p = 0.008$. There was no difference in performance between the two judgment elicitation conditions $t(108) = 0.67, p = 0.50$.

Participants were no more or less accurate on the final problem when identifying a chain structure for the second time (device 7). The most frequent error once again was mistaking the chain structure for the fully connected structure, made by 17/111 participants, although this was reduced to 11/111 when facing the chain structure again on device 7, with only a single participant making the same error twice.

Average edit distance between sequential judgments about the same device was signifi-

[17] Fifty-eight participants' final w_B judgments were incorrectly stored, so the N for w_B judgments was 53 rather than 111.

cantly increased by removing the record of previous judgments between trials, going from .73 in the *remain* condition to 1.0 in the *disappear* condition $t(109) = 3.5, p < .001$. Edit distances even in the *disappear* condition were still significantly lower than those predicted by UU, SU or SS posterior or random sampling (all p's $< .001$). As in Experiment 1 there was a strong negative relationship between number of edits and performance $F(1, 109) = 102, \beta = 6.4, \eta^2 = .48, p < .001$. The edit-distance–performance relationship interacted weakly with condition $t(108) = 1.9, \beta = 1.3, p = .049$ becoming stronger in the *disappear* condition. Again, the 71 participants who scored significantly above chance ($\frac{12}{21}$ or higher by χ^2 test) had lower edit distances of 0.66 ± 0.29 than the remaining 40 participants' 1.3 ± 0.44.

Additional measures

Participants' edge confidence judgments increased significantly over trials $\chi^2(1) = 2060, \beta = .04, SE = .0008, p < .001$, going from $.57 \pm .20$ on the first trial to $.78 \pm .19$ by the final trial. The probability of changing an edge at the next time point was weakly inversely related to the learners' reported confidence in it $\chi^2(1) = 67, \beta = -.03, SE = .004, p < .001$. Reported edge confidences were correlated with both the conditional probability of the edge states given the the rest of the current model r^{cond} =.20 and the *marginal* probability of the edge-state in the full posterior under the UU prior r^{mar} =.17 but these correlations did not differ significantly.

As predicted, reported outcome predictions were more closely related to the predictive distribution under the participants' latest structure judgment b^{t-1}: $\chi^2(1) = 1044, \beta = .35, SE = .010, p < .001$ than marginalized over the full posterior $\chi^2(1) = 580, \beta = .29, SE = .012, p < .001$. The latest-structure to prediction relationship was significantly stronger than the marginal posterior to prediction relationship by Cox test $Z = 10.9, p < .001$.

Interventions

The overall distribution of intervention choices was broadly similar to the other Experiments (Figure 11). "One-on" interventions were the most frequently chosen, making up 39% of selections. However, unlike the previous Experiments, and consistent with *edge* focused learning, the constrained "one-on one-off" interventions were almost as common as single "one-on" interventions, making up 38% of tests compared to 12% across 3-variable problems in Experiment 1. The intervention selections and informativeness of intervention sequences were not closely consistent with global expected information, nor any single type of local focus, but could again be consistent with a mixture of local *effect* focused, *edge* focused and *confirmation* focused queries.

Free explanations

For device 7, participants gave free explanations for their intervention choices on each of their six tests. The overall distribution of intervention choices did not differ significantly from the original presentation of the chain (device 3) $\chi^2 = 31, p = 0.21$ suggesting that the different response format did not affect the intervention choices that participants made. In order to assess what the explanations tell us about participants' intervention choices, we asked two independent coders to categorize the free responses into 8 categories. The categories were cho-

sen in a partly data-driven, partly hypothesis-driven way: 1. An initial set of categories were selected, with the goal of distinguishing the approximations introduced in *A local uncertainty schema* from global strategies like uncertainty sampling or expected information maximization. 2. A subset of the data was then checked and the categories were refined to better delineate their responses with minimal membership ambiguity.

The eight resulting categories were:

1. The participant just wanted to learn about one specific connection. [Corresponding to *edge focused* testing]
2. The participant wanted to learn about two specific connections.
3. The participant wanted to learn about all three connections. [Corresponding to *globally focused* testing]
4. The participant wanted to learn what a particular component can affect but did not mention a specific pattern of connections. [Corresponding to *effect focused* testing]
5. The participant wanted to test / check / confirm their current hypothesis. [Corresponding to *confirmatory* testing]
6. The participant wanted to learn about the randomness in the system (as opposed to the location of the connections). [Corresponding to a focus on learning about noise rather than structure]
7. The participant chose randomly / by mistake / to use up unwanted tests / they say they did not understand what they are doing /it is clear they were not engaging with the task.
8. The participant's explanation was complex / underspecified / did not seem to fall in any of the above categories.

A supplementary file (available at `http://www.ucl.ac.uk/lagnado-lab/el/ns_sup`) contains all the materials given to coders and the full set of participant responses. Coders were permitted to assign more than one category per response, but had to select a primary category. When the category referred to particular component label(s), the rater would record these, and when it referred to a specific connection they would record which direction (if specified) and the components involved. These details will be used to facilitate a quantitative comparison between participants' explanations and our model fits in the next Section. Raters normally just selected one category per response, selecting additional categories on only 8% of trials. Inter-rater agreement on the primary category was 0.73, and Cohen's $\kappa = 0.64 \pm 0.04$, both higher than their respective heuristic criteria for adequacy of 0.7 and 0.6 (Krippendorff, 2012; Landis & Koch, 1977).

Figure 12, shows the proportion of responses in the different categories across the six trials. On the first trial participants were most likely to be categorized as 4. – focused on identifying what a particular variable could effect. On subsequent trials they most frequently categorized as 1. – focusing on learning about a specific connection. Toward the end, explanations became more diverse and were increasingly categorized as 5. – confirmatory testing or 6. learning about the noise in the system. Individuals almost always gave a range of different explanations across their six tests, falling under 3.0 ± 0.99 different categories on average, with only $5/111$

participants providing explanations from the same category all six times (3 all-fours, 1 all-threes. and 1 all-eights).

Explanation type was predictive of performance $F(8, 657) = 13.75, \eta^2 = 0.14, p < 0.001$. Taking category 7 – unprincipled or random intervening – as the reference category with low average performance of 10.2 points out of a possible 21, categories 1,2,4,5, and 6 were all associated significantly higher final scores $[14.5, 12.9, 13.9, 13.9, 13.9]$ points, all p's < 0.001.

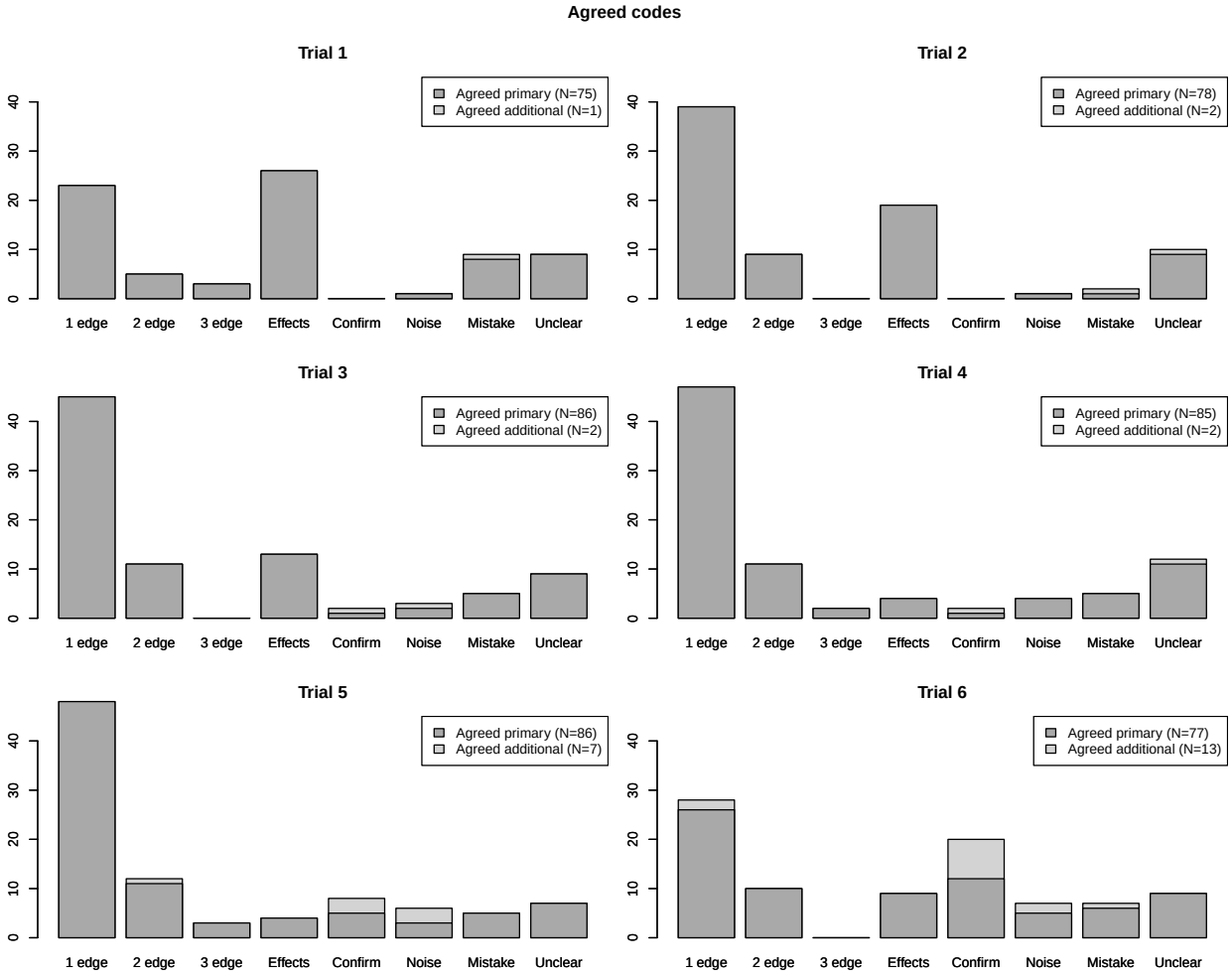

Figure 12: Experiment 2: Unknown strengths; free explanations for interventions agreed codes over the six tests in problem 7.

Summary of Experiments

In all these experiments, participants were clearly able to generate plausible causal models but also did so suboptimally. Averaged across participants, final model judgments resembled the posterior over models (e.g. Figures 4c and 9c), however individuals' trajectories typically exhibited strong sequential dependence, with the probability of moving to a new model decreasing with its edit distance from the previous model. This is consistent with our hypothesis that individuals normally maintain a single hypothesis and update it piece by piece. As found

in previous research, participants were worst at separating the direct and indirect causes in the chain (3; 8) and fully-connected (5; 10) models. A closer look at participants' intervention choices suggests that this was due to a common failure to generate the constrained interventions, such as $\mathrm{Do}[x = 1, y = 0]$, necessary to disambiguate these options. The *simple endorser* model predicts this error by proposing that people ignore the dependencies between the different edges. Our framework provides a more nuanced explanation. Whether a learner will correctly disambiguate these options depends on whether they focus on $x - z$ before or after having inferred $x \rightarrow y$ and $y \rightarrow z$. If the consider $x - z$ after, then they will tend to fix y "off", realizing it is necessary to prevent the indirect path from confounding the outcome of their test. However, if they have no connection marked from x to y or from y to z, they will not expect this confounding activation and so have no motivation to fix y "off" when testing $x - z$.

Participants' overall distributions of intervention selections resembled a mixture of *edge*, *effect* and *confirmation* focused testing, but their distributions of choices were relatively invariant across conditions and trials while the efficient learners' were much more dynamic. Comparison with the final global information gathered revealed that they did not select which variables to target particularly efficiently, leading to a considerable discrepancy between the total information gathered by participants compared to an ideal active learner. However, participants also displayed hints of adaptation of strategy over the trials: with a preference for confirmatory testing, being more likely to fix variables "on" when they had children according to their latest hypothesis b^{t-1}, and displaying a modest shift toward more constrained interventions in later trials.

In Experiment 2 we saw that people were able to identify causal structure effectively without specific parameter knowledge. Comparing a range of plausible prior assumptions about edge reliability w_S and the level of background noise w_B yielded little difference in judgment or intervention choice predictions. Participants' overall judgment accuracy was not affected by the *remain/disappear* reporting condition, but this did affect sequential dependence, especially for lower performers who may have often forgotten their previous judgment when making their next one. The idea, common to the three judgment rules we consider, that people represent one model at a time was also supported by the additional measures elicited from participants during the task. With a single hypothesis rather than distributional beliefs, intervention outcome predictions could only be generated by the current hypothesis rather than averaged and weighted over all possible models. Consistent with this idea, we found participants' expectation judgments were more in line with their current hypothesis than the marginal likelihoods, although we note that these measures were quite noisy and the effects quite small.

Modeling individual behavior

Across all three examined experiments we found a qualitative correspondence, both between our *Neurath's ship* simulations and participants' judgments, and between the two stage local intervention schema and participants interventions. However, both *simple endorsement* and *win-stay, lose-sample* also appeared to do a good job of capturing qualitative judgment patterns. In order to validate quantitatively which of these models better describes participants'

behavior, we fit the models to the data and assessed their competence relative also to *win-stay, lose-sample* and *simple endorsement*. By fitting the models separately to individual participants we also assessed individual differences in learning behavior, and thus gained a finer-grained picture of the processes involved.

Judgments

Models

We compared six models to participants judgments, the three process models we considered in the experiment *Neurath's ship* (NS), *simple endorser* (SE), *win-stay,lose-sample* (WSLS), alonside an efficient Bayesian learner (*Rational*) and two null models *Baseline* and NS-RE.

For NS, we fit three parameters:

1. An average search length parameter λ controlling the probability of searching for different lengths k on each belief update.
2. A search behavior parameter ω controlling how strongly the learner moves toward the more likely state for an edge when updating it (recalling that $\omega = 1$ leads to probability matching, while $\omega = \infty$ leads to deterministic hill climbing and $\omega = 0$ to making random local edits).
3. A lapse parameter ϵ controlling a mixture between the model predictions and a uniform distribution.

Including the last parameter into equation 5, this resulted in the following equation

$$P(b^t = m|\mathcal{D}_r^t, \mathcal{C}_r^t, \mathbf{w}, b^{t-1}, \omega, \lambda) = (1-\epsilon)\sum_0^\infty \frac{\lambda^k e^{-\lambda}}{k!}[(R_t^\omega)^k]_{b^{t-1}m} + \epsilon \text{Unif}(M) \tag{13}$$

where R is a Markov matrix expressing the options for local improvement.

We operationalised the *Simple endorser* (SE) (Bramley, Lagnado, & Speekenbrink, 2015) with two parameters. One is the probability $\rho \in [0, 1]$ with which the belief state is updated from b^{t-1} include extra edges from any currently fixed "on" node(s) to any activated nodes and to exclude edges from any currently fixed "on" node(s) to any non-activated nodes (we write b_{+SE}^{t-1}). With the complementary probability $1 - \rho$, it stays the same as b^{t-1}. As with the NS model we also included a lapse parameter mixing in a probability of choosing something at random, giving

$$P(b^t = m|\mathbf{d}, \mathbf{w}) = (1-\epsilon)(\rho\, b_{+SE}^{t-1} + (1-\rho)b^{t-1}) + \epsilon Unif(M) \tag{14}$$

Win-stay, lose-sample (WSLS) (Bonawitz et al., 2014) predicts that participants stick with their current model b^{t-1} with probability $p(\mathbf{d}^t|b^{t-1}, \mathbf{w}, c^t)$ or else draw a sample from the full posterior with probability $1 - p(\mathbf{d}^t|b^{t-1}, \mathbf{w}, c^t)$. The fitted version of this model had a single lapse parameter ϵ giving

$$P(b^t = m|D^t, \mathbf{w}) = (1-\epsilon)\Big((1-P(\mathbf{d}^t|b^{t-1}, \mathbf{w}, c^t))P(M|D^t, \mathbf{w})_t + P(\mathbf{d}^t|b^{t-1}, \mathbf{w}, c^t)[m = b^{t-1}]\Big) + \epsilon \text{Unif}(M) \tag{15}$$

The final model, *Rational* was a variant of the Bayes-optimal observer (Section 2) that attempted to select the maximum a posteriori causal structure $\max P(M|D^t, \mathbf{w}; C^t)$ with each judgment, with a soft maximization (Luce, 1959) governed by inverse temperature parameter θ and a lapse parameter ϵ. For this, we considered

$$P(b^t = m|D^t, \mathbf{w}) = (1-\epsilon)\frac{\exp(P(M|D^t, \mathbf{w})_t\theta)}{\sum_{m' \in M}\exp(P(m'|D^t, \mathbf{w})_t\theta)} + \epsilon \text{Unif}(M) \tag{16}$$

Baseline is a parameter-free baseline that assumes each judgment to be a random draw from all possible causal models

$$p(b^t = m) = \text{Unif}(M) \tag{17}$$

(leading to a probability of approximately $\frac{1}{3}$ for each edge).

One concern with this baseline is that judgments might exhibit sequential dependence yet be unrelated to data $\mathcal{D}_r^t$. Therefore we also considered a baseline variant of the NS model in which the search behavior parameter ω was fixed to 0, resulting in a (R)andom (E)dit model (NS-RE) that walks randomly around the hypothesis space for k steps on each update. For this model, small k simply denotes more inertia.

Each of these belief models output a likelihood based on the probability that that the model assigns to a belief of b^t, given the most recent outcome $\mathbf{d}^t$ (SE), outcomes since the last belief change $\mathcal{D}_r^t$ (NS), or all outcomes D^t (WSLS, *Rational*), and the most recent judgment b^{t-1}. Because the choice of prior for Experiment 2 made negligible difference to our results, we only report models assuming uniform (UU) priors on $\mathbf{w}$. For Experiment 2, we also marginalized over the unknown values of $\mathbf{w}$ rather than conditioning as in the other experiments as detailed in Appendix B.

Evaluation

To compare these models quantitatively, we used maximum likelihood optimization as implemented by R's `optim` function to fit the model separately to each of the 370 participants across all three experiments.[18] We used Bayesian Information Criterion (BIC, Schwarz, 1978) to compare the models while accommodating their differing numbers of parameters. *Baseline* acts as the null model for computing BICs and pseudo-R^2s (Dobson, 2010) for the other models. In Bramley, Lagnado, and Speekenbrink (2015) participants were not forced to select something for each edge immediately, although once they did so they could not return to "unspecified", and they could also respond with cyclic causal model if they wanted. Therefore, we fit only the 75% of tests where the participants report a fully specified non-cyclic belief, taking the b^{t-1} to be the unconnected model on the first fully specified judgment, as we do with b^0 in the other Experiments. Recalculating the transition probabilities on the fly in the optimization of ω was infeasibly computationally intensive for the four-variable problems. So for Experiment 1 we first fit all three parameters to the three-variable problems only, then used the best fitting ω parameters from this fit when fitting the λ and ϵ on the full data. In Bramley, Lagnado, and Speekenbrink (2015) and Experiment 2 we were able to fit all three parameters.

[18] In Appendix B we provide additional detail on how the models were fit.

Results and discussion

Table 3 details the results of the model fits to all experiments. Summed across all participants, NS has the lowest total BIC (93381) with the SE in second place (94326), followed by WSLS with (97643), then NS-RE (101837), *Rational* (1207209) and finally *Baseline* with (149313). NS was also the best fitting model for Bramley, Lagnado, and Speekenbrink (2015) and Experiment 1, with SE winning in Experiment 2. Thus, all three heuristics substantially beat an exact Bayesian inference account of causal judgment here, but *Neurath's ship*, with its ability to capture a graded dependence on prior beliefs, outperformed WSLS substantially, and the heuristic SE to a lesser degree. In terms of number of individuals best fit, Table 3 shows a broad spread across models: WSLS – 102, NS – 85, SE – 80, NS-RE – 70, *Rational* – 28 , *Baseline* – 4.

Table 3: Belief Model Fits.

Bramley, Lagnado, and Speekenbrink (2015)															
Model	λ M	λ SD	ω M	ω SD	θ M	θ SD	ρ M	ρ SD	ϵ M	ϵ SD	N fit	M acc	logL	R^2	BIC
Baseline											1	0.27	-17836	0	35672
NS-RE	0.17	51.5							0.21	0.19	32	0.36	-9379	0.49	19762
SE							0.18	0.18	0.21	0.24	13	0.60	-8819	0.53	18642
WSLS									0.14	0.28	56	0.85	-9117	0.52	18736
NS	**1.20**	**100.1**	**6**	**243.8**					**0.05**	**0.20**	**27**	**0.66**	**-8197**	**0.56**	**17901**
Rational					5	124.8			0.00	0.39	10	0.93	-12089	0.36	25182
Exp 1: Learning larger models															
Model	λ M	λ SD	ω M	ω SD	θ M	θ SD	ρ M	ρ SD	ϵ M	ϵ SD	N fit	M acc	logL	R^2	BIC
Baseline											1	0.33	-41814	0	83628
NS-RE	1.23	1.5							0.25	0.28	33	0.55	-29235	0.30	59490
SE							0.51	0.32	0.56	0.29	30	0.49	-27736	0.34	56492
WSLS									0.32	0.34	29	0.52	-28772	0.31	58053
NS	**1.63**	**1.7**	**4**	**241.0**					**0.20**	**0.29**	**27**	**0.56**	**-27234**	**0.35**	**55896**
Rational					13	143.2			0.14	0.36	0		-36362	0.13	73743
Exp 2: Unknown strengths															
Model	λ M	λ SD	ω M	ω SD	θ M	θ SD	ρ M	ρ SD	ϵ M	ϵ SD	N fit rem/diss	M acc	logL	R^2	BIC rem/diss
Baseline											0/2	0.21	-15006	0	14330/15682
NS-RE	0.9	3							0.10	0.29	3/2	0.36	-10877	0.28	10050/12535
SE							**0.80**	**0.30**	**0.43**	**0.27**	**21/16**	**0.75**	**-9181**	**0.39**	**8257/10936**
WSLS										0.31	9/8	0.50	-10220	0.32	9402/11452
NS	1.8	121	9.2	247					0.03	0.20	13/18	0.55	-9170	0.39	8620/10964
Rational					31.5	282				0.26	7/12	0.70	-10482	0.30	10116/11678

Note: Columns: M = median estimated parameters across all participants, SD = standard deviation of parameter estimate across all participants, N fit = number of participants best fit by each model, M acc = average proportion of edges identified correctly by participants best fit by this model, LogL = total log likelihood of model over all participants, R^2 = median McFadden's pseudo-R^2 across all participants, BIC = aggregate Bayesian information criterion across all participants. For Exp 2, rem = remain condition, diss =disappear condition. Best fitting model denoted with boldface.

The diversity of individual fits across strategies raises the question of the identifiability of the different models. To assess how reliably genuine followers of the different proposed strategies would be identified by our modelling procedure, we simulated participants using the fitted parameters for each model for each of the actual participants in all three examined experiments. We then fit all six models to these simulated participants report the rates at which

simulations are best-captured by each model. Table 5 in the Appendix provides the complete results for this recovery analysis. Overall, the generating model was recovered 74% of the time for Bramley, Lagnado, and Speekenbrink (2015), 82% for Experiment 1 and 75% for Experiment 2 (chance would be 17%). In all three experiments, data generated by *Baseline, WSLS* and SE was nearly always correctly recaptured, indicating that we can treat cases where participants are well described by these models as genuine. Additionally NS almost never captured data generated by any of the other models, providing reassurance that NS is not simply fitting participants who are doing something more in line with SE or WSLS. However, data actually generated by NS was frequently recaptured by the NS-RE (random edit) null model that makes NS-style local edits but does not preferentially approach more likely models. This was true in the majority of cases in Bramley, Lagnado, and Speekenbrink (2015) and Experiment 2. Some of the cases where NS-RE captures NS-generated simulations are based on participants who were better described by NS-RE in the first place (e.g. whose search behaviour was too random to justify ω's inclusion). We find a similar effect whereby simulated *Rational* participants with relatively low θs or high ϵs are more parsimoniously described by *Baseline*. This is supported by looking at the more complex four variable problems in Experiment 1, NS simulations were identified the majority of the time, and when restricted to simulations based on parameters from participants who were actually best described by NS, 24/27 were recovered successfully. Thus it is plausible that some of the 70 NR-RE participants were infact doing something more in line with NS. There is a suggestion of this in Experiment 1, where the mean accuracy of the NS-RE participants is commesurate with SE, WSLS and NS.

The performance of a handful of participants – 10 in Bramley, Lagnado, and Speekenbrink (2015), and 19 in Experiment 2 – were best fit by the *Rational* model, which has one fewer parameter than NS. Naturally, these participants performed particularly well, scoring near ceiling in Bramley, Lagnado, and Speekenbrink (2015) (identifying 14.0 of the 15 connections) and as high as the ideal learning simulations in Experiment 2 – 14.7/21 compared to an average of 15.5 for perfect Bayesian integration. This, along with the lower recovery rates for these experiments, suggests that their design – both being motivated primarily to look closely at intervention choice – may not have been difficult enough to separate the process from the normative predictions about the judgments.

Figures 13a and b show the range of the fitted λ and ω parameters under NS. In line with our predictions, participants' average fitted search lengths (λ) were mainly small, with medians between 1 and 2 in all three experiments.[19] Because this parameter merely encodes a participant's average search length this means that the same participant would sometimes not search at all, staying exactly where they are ($k = 0$), or might also sometimes search much longer (e.g. $k \gg \lambda$). The median fitted ωs of 6, 4 and 9.2 across the three experiments are suggestive of moderate hill-climbing. A substantial number of participants had very large values of ω indicative of near-deterministic hill climbing. We discuss this trade-off further in the General Discussion. However, note that we were only able to fit these values to the easier three variable problems. It might be that the largest values would have been tempered if they could

[19] A few participants made judgments that were sequentially anti-correlated leading to λ parameters at the limit of the optimization routine's precision and correspondingly large standard deviations in Experiments 1 and 2.

have been fit to the four variable problems as well.

Interventions

Models

We compared our local model of intervention choice (Section 4) to a globally-focused and a baseline model. Each intervention model output a likelihood for an intervention choice of $\mathbf{c}^t$, depending on $\mathcal{D}_r^t, \mathcal{C}_r^t$ and b^{t-1}.

We compared the overall distribution of participants' intervention selections and final performance with *edge* focused, *effect* focused and *confirmation* focused tests. We found that none of these models alone closely resembled participants' response patterns, but overall distributions were consistent with a mixture of different types of local tests. This was also supported by the the free-response coding in Experiment 2, showing that participants would typically report targeting a mixture of specific edges, effects of specific variables and confirming the current hypothesis. Therefore, we considered four locally driven intervention selection models, one for each of the three foci, plus a mixture.

For the *edge* model, the possible foci $\mathcal{L}$ included the 3 (or 6) edges in the model. For the *effect* model, it comprised the 3 components (or 4 in the 4-variable case). The confirmation model always had the same focus – comparing b^t to null b^0 of no connectivity. The *mixed* model contained all 7 (or 11) foci. As in Equations 8 and 9 in Section 4, each model would first compute a soft-max probability of choosing each possible focus $\mathbf{l}^t \in \mathcal{L}$. Within each chosen focus it would also calculate the soft-max probability of selecting each intervention, governed by another inverse temperature parameter $\eta \in [0, \infty]$. The total likelihood of the next intervention choice was thus a soft-maximization-weighted average of choice probabilities across possible focuses

$$P(\mathbf{c}^t|\eta, \rho, \mathcal{D}_r^t, b^{t-1}, \mathbf{w}) = \sum_{\mathbf{l}\in L} P(\mathbf{c}|\mathbf{l}, \eta, b^{t-1}, \mathbf{w}) \frac{\exp(H(\mathbf{l}|\mathcal{D}_r^t, b^{t-1}, \mathbf{w}; \mathcal{C}_r^t)\rho)}{\sum_{\mathbf{l}'\in\mathcal{L}} \exp(H(\mathbf{l}'|\mathcal{D}_r^t, b^{t-1}, \mathbf{w}; \mathcal{C}_r^t)\rho)} \tag{18}$$

where

$$P(\mathbf{c}|\mathbf{l}, \eta, b^{t-1}, \mathbf{w}) = \frac{\exp\left(\mathbb{E}_{\mathbf{d}\in\mathcal{D}_{\mathbf{c}^t}}\left[\Delta H(\mathbf{l}|\mathbf{d}, b^{t-1}, \mathbf{w}; \mathbf{c}\right]\eta\right)}{\sum_{\mathbf{c}'\in\mathcal{C}} \exp\left(\mathbb{E}_{\mathbf{d}\in\mathcal{D}_{\mathbf{c}}}\left[\Delta H(\mathbf{l}|\mathbf{d}, b^{t-1}, \mathbf{w}; \mathbf{c}'\right]\eta\right)} \tag{19}$$

Positive values of $\rho \in [-\infty, \infty]$ encode a preference for focusing on areas where the learner should be most uncertain, $\rho = 0$ encodes random selection of local focus, and negative ρ encodes a preference for focusing on areas where the learner should be most certain.

For comparison, *Baseline* is a parameter-free model that assumed each intervention was a random draw from all possible interventions

$$P(\mathbf{c}^t) = \mathrm{Unif}(\mathcal{C}) \tag{20}$$

Global is a variant of the globally efficient intervention selection (Section 2) that attempted to select the globally most informative greedy test $\arg\max_{\mathbf{c}\in\mathcal{C}} \mathbb{E}_{\mathbf{d}\in\mathcal{D}_{\mathbf{c}}}\left[\Delta H(M|\mathbf{d}, D^{t-1}, \mathbf{w}; C^{t-1}, \mathbf{c})\right]$. It has one inverse temperature parameter $\theta \in [0, \infty]$ governing soft maximization (Luce, 1959)

over the global expected information gains. For this, we considered

$$P(\mathbf{c}^t|D^{t-1},\mathbf{w};C^{t-1}) = \frac{\exp(\mathbb{E}_{\mathbf{d}\in\mathcal{D}_{\mathbf{c}^t}}\left[\Delta H(M|\mathbf{d},D^{t-1},\mathbf{w};\mathbf{c}^t)\right]\theta)}{\sum_{\mathbf{c}\in\mathcal{C}}\exp(\mathbb{E}_{\mathbf{d}\in\mathcal{D}_{\mathbf{c}}}\left[\Delta H(M|\mathbf{d},D^{t-1},\mathbf{w};\mathbf{c})\right]\theta)} \tag{21}$$

As with the belief modeling, for Experiment 2 we marginalized over the the unknown values of $\mathbf{w}$ rather than conditioning as in Experiments 1-2 as detailed in Appendix B.

Evaluation

All six models were fit to the data from all three experiments in the same way as the belief models. The results are detailed in Table 4.

Additionally, to compare model predictions of local focus choice $\mathbf{l}^t$ to participants' self reports in problem 7 in Experiment 2, we computed the likelihood of each local focus prediction on each test. This was done by calculating $P(\mathbf{c}|\mathbf{l},\eta,b^{t-1},\mathbf{w})$ for each of the local foci we considered, using a fixed common $\eta = 20$ to capture strong but non deterministic preference for the most useful intervention(s). For each data point c^t, we then calculated which l^t assigned the most probability to c^t the intervention actually chosen by the participant. Figure 14 plots the most likely focus of participants' intervention choices in the final problem against the code assigned to their free responses.

Results and discussion

Table 4: Intervention Model Fits

	Bramley, Lagnado, and Speekenbrink (2015)										
	Model	η M	η SD	ρ M	ρ SD	θ M	θ SD	N fit	M acc	logL	R^2	BIC
1	Baseline							3	0.38	-18262		36524
3	Edges	10.4	26	0.9	80			9	0.68	-14222	0.23	29338
2	**Effects**	**7.7**	**16**	**0.5**	**5**			**82**	**0.71**	**-10701**	**0.41**	**22296**
4	Confirmatory	3.9	74					9	0.43	-15368	0.14	31182
5	Mixed	15.1	152	0.7	33			32	0.66	-11145	0.39	23185
6	Global					6.1	4.1	4	0.85	-15619	0.14	31686
	Exp 1: Learning larger models											
	Model	η M	η SD	ρ M	ρ SD	θ M	θ SD	N fit	M acc	logL	R^2	BIC
1	Baseline							18	0.35	-32958		65917
2	Edges	9.3	138	25.9	1012			2	0.37	-28588	0.13	58196
3	Effects	6.2	39	1.5	6			31	0.59	-24213	0.27	49445
4	Confirmatory	3.8	8					24	0.49	-28721	0.13	57951
5	**Mixed**	**8.8**	**139**	**17.6**	**414**			**27**	**0.61**	**-23944**	**0.27**	**48907**
6	Global					4.9	4	18	0.66	-26652	0.19	53813
	Exp 2: Unknown strengths											
	Model	η M	η SD	ρ M	ρ SD	θ M	θ SD	N fit	M acc	logL	R^2	BIC
1	Baseline							14	0.35	-15365		30730
2	Edges	4.0	8	2.7	77			24	0.74	-13010	0.18	26850
3	Effects	3.2	9	2.9	218			7	0.52	-12992	0.14	26815
4	Confirmatory	2.5	14					12	0.46	-14180	0.04	28776
5	**Mixed**	**3.9**	**9**	**5.4**	**285**			**15**	**0.63**	**-12550**	**0.17**	**25931**
6	Global	3.0	8					39	0.70	-12850	0.14	26114

Note: Columns as in belief model (Table 3).

The *mixed* local focus model was the best fitting model over the three experiments with the

lowest total BIC of 97757, followed by *effects* then by the *global* focused model, then by *edges* and finally by *confirmation* and then baseline. However, there was a great deal of individual variation, suggesting that a single model does not capture the population well. More participants were best described by an *effects* focus (121) than a *mixed* focus (77), but each model received some support, with 58, 43, 36 and 35 individuals best fit by *global*, *confirmation* and *edge* focused and *baseline* models respectively. Additionally, the *effect* focus was was the best fitting model overall in Bramley, Lagnado, and Speekenbrink (2015) where there was a strong tendency for participants to fix a single variable on at a time.

As Table 4 shows, *mixed* was the best overall fitting model for Experiments 1 and 2, and the majority of participants 277/370 were fit by one of the local uncertainty driven models. Furthermore, Figure 14 shows that for effect and edge queries, there was a strong correspondence between the most likely choice of focus l on Experiment 2 problem 7 and the coded explanation of that intervention's goal. This was not the case for tests where explanations were categorized as *confirmatory*. These were most frequently best described as effect focused tests of the root variable of the true model (labeled "x" in the plots).

As with the case of judgments, a moderate number of chance-level performing participants (35/370) were best described by the *Baseline* model. However, 58 participants across the three experiments were better described by the *Global*ly efficient testing model than any local testing models. However, these were not the highest performing participants in Experiment 2, with lower average scores than those described by the *edge* focused model. This suggests that we do not yet have a good model of these participants' choices.

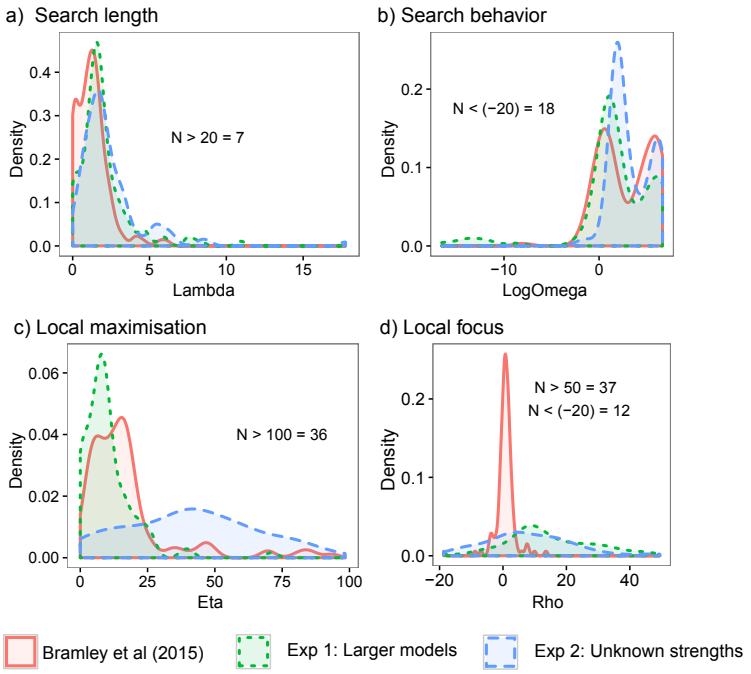

Figure 13: Gaussian kernel densities over fitted model parameters for all participants. a) Search length λ and b) log search behavior $\log(\omega)$ according to *Neurath's ship* belief update model. c) Local maximization parameter η and d) local focus choice parameter ρ under *mixed* local uncertainty based model. Since we report all participants fits, there are some extreme values — poorly described by either model — that are not plotted. Annotations give the number of parameters above and below the range plotted.

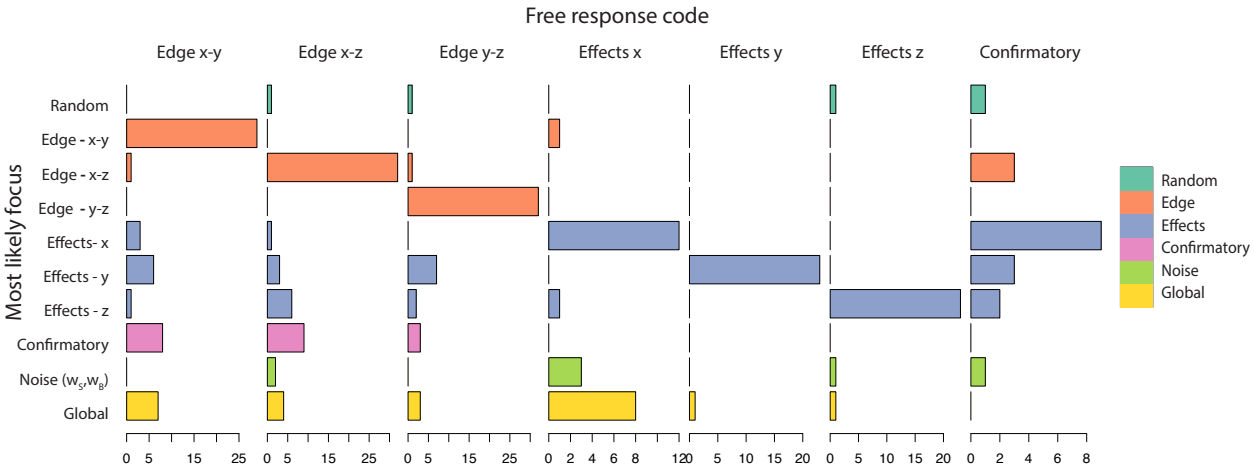

Figure 14: Model and free response correspondence. Each plot is for trials assigned a particular free response code, each bar is for the number of trials for which that local focus was most likely given the intervention choice. *Effect* and *edge* coded queries were also diagnosed as such by the model fitting while *confirmatory* coded queries were most likely to be diagnosed as querying the effects of the root node(s) in the true model which always was (or included) x.

General Discussion

Actively learning causal models is key to higher-level cognition and yet is radically intractable. We explored how people manage to identify causal models despite their limited computational resources. In three experiments, we found that participants' judgments some-

what reflected the true posterior, while exhibiting sequential dependencies. Further, participants' choices of interventions reflected average expected information, but were insufficiently reactive to the evidence that had already been observed and were consistent with being locally focused.

We could capture participants' judgment patterns by assuming that the they maintained a single causal model rather than a full distribution. We proposed that participants considered local changes to improve the ability of their single model to explain the latest data and compared this account to two other proposals, one based on the idea that participants occasionally resample from the full posterior, and the other, a heuristic based on ignoring the possibility of indirect effects. While our *Neurath's ship* proposal fit best overall, all three proposals had merit, with *simple endorsement* winning out in Experiment 2 and more individuals better fit by *win-stay lose-sample*.

We captured participants' interventions by assuming they focused stochastically on different local aspects of the overall uncertainty and tried to resolve these, leading to behavior that was comparatively invariant to the prior. Our modelling suggested a broad spread of local focuses both between and within participants.

By casting our modeling in the language of machine learning, we were able to make strong connections between our *Neurath's ship* model and established techniques for approximating distributions–sequential Monte-Carlo particle filtering and MCMC (specifically Gibbs) sampling. Likewise we were able to explicate intervention selections using the language of expected uncertainty reduction but relaxing the assumption that the goal was the global uncertainty in the full distribution. The combination of a single hypothesis (particle) and a Gibbs-esque search, nicely reflects the *Neurath's ship* intuition that theory change is necessarily piecemeal and that changes are evaluated against the backdrop of the rest of the existing theory.

Limitations of Neurath's ship

Like any theory, Neurath's ship was evaluated against a backdrop of a number of assumptions. We discuss some of these here.

Measurement effects

In order to explore incremental belief change it was necessary to elicit multiple judgments and to make two strong assumptions: (1) that these judgments reflected participants' true and latest beliefs; and (2) that the repeated elicitations did not fundamentally alter learning processes. To mitigate problems of these, we both incentivised participants to draw their best and latest guess at every time point during the tasks, and examined different reporting conditions to explore the influence of the elicitations on the learning process.

In Bramley, Lagnado, and Speekenbrink (2015) and the *remain* condition in Experiment 2, participants could leave parts of their hypothesis untouched if the did not want to change them. This had the strength of being minimally invasive; it did not push the learner to reconsider an edge that they would otherwise not have done merely because they have been asked about it again. However this came at the cost of conflating genuine incremental change in the learner's psychological representation with response laziness. To assuage this concern,

in Experiment 1 and Experiment 2 *disappear*, we removed the participants' previous judgment after they had seen the outcome of the subsequent intervention, meaning that they would have to remember and re-report any edges they had previously judged (and not yet reconsidered). The slight reduction of dependence between *remain* and *disappear* conditions in Experiment 2, is consistent with the idea that being forced to re-report edges made it more likely that they would be reconsidered and potentially changed.

The *Neurath's ship* approach is related to anchor-and-adjust models (Einhorn & Hogarth, 1986; Petrov & Anderson, 2005) of sequential magnitude estimation. Hogarth and Einhorn found that when mean estimates are repeatedly elicited from participants as they see a sequence of numbers, the sequence of responses can be captured by a process whereby one stores a single value and adjusts it a portion of the way toward each new observed value. When judgments were elicited at the end of the sequence, participants behaved more like they had stored a subset of the values and averaged them at the end. In the same way, we can think of *Neurath's ship* as a process in which the current model acts as an anchor, and adjustments are made toward new data as it is observed. However, the higher complexity of causal inference, and the greater storage requirements for the individual episodes will presumably lead to greater pressure to use a sequential strategy rather than store. Arguably, step-by-step elicitation is a closer analogue to real-world causal inference than end-of-sequence because causal beliefs are presumably in frequent use while learning instances may be spread out, with no clear start or end.

Acyclicity

We adopted the directed acyclic graph as our model of causal representation here because it is a standard approach in the literature and is mathematically convenient. Furthermore, cyclic graphs were quite rare choices in Bramley, Lagnado, and Speekenbrink (2015) (where participants were permitted to draw them). Thus, we simply opted to to rule them out in the instructions in later experiments.

However, in tasks where people draw causal models of real-world phenomena, they often draw cyclic or reciprocal relationships (Kim & Ahn, 2002; Nikolic & Lagnado, 2015), and many real world processes are characterized by bidirectional causality, such as supply and demand in economics or homeostasis in biological systems. There are various ways to represent dynamic systems. One proposal is the dynamic Bayesian network (Dean & Kanazawa, 1989), which can be "unfolded" to form regular acyclic network with causal influences passing forward through time. Another is the chain graph (Lauritzen & Richardson, 2002), in which undirected edges are mixed with directed edges and used to model the equilibria of the cyclic parts of the system.

Exploring these structures would require a change in the the semantics of the experiment so that people could understand what they were reporting in the presence of dynamical interactions. However, given this, NS would offer a way of performing sequential, on-line, inference for such structures, using standard likelihood calculations for dynamic Bayes nets and chain graphs.

Evaluation of evidence

Another pragmatic limitation of the current modeling was the assumption of the noisy-OR functional form for the true underlying causal models. While we did take care to train participants on the sources of noise in both Experiments and the exact values in in Bramley, Lagnado, and Speekenbrink (2015), our own past work suggests that people may have simpler ways of evaluating the how likely models would be to produce different patterns – for example, in Bramley, Dayan, and Lagnado (2015), we found participants' judgments could be captured by assuming they lumped sources of noise together and just counted the number of surprising outcomes under each model.

One possibility is that people actually formed likelihood estimates through simulation with an internal causal model. For instance, one might perform a mental intervention, activating a component of one's own internal causal model and keeping track of where the activation propagates. By simulating multiple times, a learner could estimate the likelihood of different outcomes under their current model (Hamrick, Smith, Griffiths, & Vul, 2015), and by simulating under variations of the model, the learner could compare likelihoods generated on the fly. This simulation-based view provides a possible explanation for why participants more readily accommodated internal noise w_S than background noise w_B. The former can be "built in" to the inferred connections in their model and reveal itself in mental simulation, while w_B is more of a mathematical "catch all" for all possible influences coming from outside the variables under focus. The Neurath's ship perspective suggests that people lean on their surrounding network of assumptions about surrounding causes, controlling for these if they get in the way of local inference. By being omnipresent and affecting all the variables equally w_B was not possible to accommodate in this way.

Future experiments and modeling might relax the assumption of noisy-OR likelihoods and allow induction of more diverse functional forms, or focus on well known domains where priors can be measured before the task. Another approach might be to render the noisy-OR formalization more transparent by visualizing the sources of exogenous noise alongside the target variables, for instance displaying varying numbers of nuisance background variables on screen for different background noise conditions.

Antifoundationalism

The core of Neurath's ship is the strong assumption that people consider only a single global hypothesis and make local changes within this. This is the "antifoundationalism" captured by Duhem-Quine thesis–any local theoretical claim is necessarily supported by surrounding assumptions. However, this may be too strong for some of the easier problems we considered here where the worlds may have been small and constrained enough for some people to reason at the global level. For the three variable problems in particular, some participants may have been able to consider alternatives at the level of the whole model, and thus able to shift from common cause to chain etc with a single step.

While participants' judgments showed high sequential dependence, they did occasionally change their model abruptly. The theory of unexpected uncertainty (Yu & Dayan, 2003), and substantial work on changepoint tasks (Speekenbrink & Shanks, 2010) are associated with the

notion that people will sometimes "start over" if they are having consistently poor predictions from their existing model. This relates to the idea, in philosophy of science, of a "paradigm shift" (**?**). The current *Neurath's ship* models do not naturally capture this but accommodate occasional large jumps by assuming a variable search length (k), meaning the search will sometimes be long enough to allow the learner to move to a radically different model in a single update. However we might also extend the *Neurath's ship* framework to include a threshold on prediction accuracy below which a learner will start afresh, for example by randomly sampling a model, or sampling from a hitherto unexplored part of the space. At present this is captured by the ϵ probability of sampling a new b^t at random on a given trial (which ranged between a probability of .03 in Experiment 2 and .2 in Experiment 1).

Selective memory

We assumed that participants' judgment updates were based on the recent data $\mathcal{D}_r^t$, collected since the last time they changed their hypothesis. This is quite frugal in the current context, as the learner rarely has to store more than a few tests worth of evidence. It also captures the idea of semanticization – that as one gradually absorbs episodic evidence into one's hypothesis, it becomes safe to forget it.

However, the particular choice of $\mathcal{D}_r^t$ is certainly a simplification. People may frequently remember evidence from before their latest change, and fail to store recent evidence, especially once their beliefs become settled. They might also collect summary evidence at the level of individual edges, counting how often a pairs components activate together for example, or remember evidence about some components but not others, or only store evidence when it is surprising under the current model. In order to fit the models it was necessary to make simplifying assumptions that captured some form of halfway house between remembering everything and relying entirely on your hypothesis. Future studies might probe exactly what learners can remember during and after learning to get a finer-grained understanding of the trade off between remembering evidence and absorbing it into beliefs.

Related to this, we fit a static search behavior parameter to participants, finding evidence of moderate hill climbing. However, a more realistic depiction might be something more akin to simulated annealing (Hwang, 1988). Learners might begin searching with more exploratory moves $\omega \approx 1$ so as to explore the space broadly, and transition toward hill climbing $\omega = \infty$ as they start to choose what judgment to report. Alternatively they might gradually reduce their search length k as pressure to settle on a model increases.

Alternative approximations and representations

The choice of Gibbs sampling, together with a single particle approximation, is just one of numerous possible models of structure inference. For example we found (data not shown) fairly good fits by replacing Gibbs sampling with a form of Metropolis-Hastings MCMC sampling – using an MC^3 proposal and acceptance distribution (Madigan & Raftery, 1994; Madigan, York, & Allard, 1995). The two approaches make similar behavioral predictions but differ somewhat in their internal architecture – a Metropolis-Hastings sampler would first generate a wholesale alternative to the current belief, then make an accept-reject decision about whether

to accept this alternative, while the Gibbs sampler focuses on one subpart at a time and updates this conditional on the rest. Ultimately, the Gibbs sampler did a better job, helping justify the broader ideas of locality of inference implicit in the *Neurath's ship* proposal.

An interesting alternative approach to complex model induction via local computations (Fernbach & Sloman, 2009; Waldmann et al., 2008), comes from variational Bayes (Bishop, 2006; Weierstrass, 1902). The idea behind this is that one can simplify inference by replacing an intractable distribution, here the distribution over all possible models, with a simpler one which has degrees of freedom that can be used to allow it to fit as best as possible. A common choice of simpler distribution involves factorization, with a multiplicative combination of a set of simpler parametrized distributions. Thus, for causal inference one might make a mean-field approximation (Georges, Kotliar, Krauth, & Rozenberg, 1996) and suppose the true distribution over models factorizes into independent distributions for each causal connection. Divergence between this approximation and the full model can then be minimized mathematically by updating each of the local distributions in turn (Jaakkola, 2001). This provides a different perspective on global inference based on local updates. Rather than a process of local search where only a single model is represented at any time, variational Bayes suggests people maintain many local distributions and try to minimize the inconsistencies between them. The biases induced by this process make the two approaches distinguishable in principle (Sanborn, 2015), meaning that an interesting avenue for future work may be to design experiments that distinguish between the two approaches to approximation in cognition . The truth in our case may be somewhere in between. For instance, in the current work, we assumed people were able to use recent evidence to estimate their local uncertainty conditional on the rest of the structure, and thus choose where to focus interventions. To the extent that learners really represent their beliefs with lots of local uncertainties, their representation becomes increasingly variational.

Choosing interventions aboard Neurath's ship

The largest difference in intervention choices between experiments was that in Experiment 2 constrained interventions (e.g. $\mathrm{Do}[x = 1, y = 0]$) were chosen much more frequently. One explanation for this is that participants might have been forced to focus their attention more narrowly in Experiment 2, to compensate for their additional uncertainty about the noise by using more focused testing. Another possibility is that the different subject pools drove this difference. It is possible that mTurk's older and educationally diverse participants (Experiments 1) gathered evidence differently from the young scientifically trained UCL undergraduates (Experiment 2). This might have driven the tendency toward more tightly constrained tests in in Experiment 2.

The idea that people relied on asking a mixture of different types of locally focused question, was borne out by our analysis of the coding of participants' free explanations. Explanations almost always focused on one specific aspect of the problem, most frequently on a particular causal connection, or what a particular component can affect, but also sometimes on parameter uncertainty or, on later tests, confirming their current hypothesis. Furthermore, participants almost always referred to a mix of different local query types over the course of

their six tests. The apparent shift toward confirmatory testing on the last trial is sensible, since participants they knew they would not have more tests to follow up anything new they might discover. Indeed this shift would be normative in various settings.

Subjective explanations are notoriously problematic (Ericsson & Simon, 1980, 1993; Russo, Johnson, & Stephens, 1989). Therefore, we must be careful in interpreting these results. One common issue is that eliciting responses concurrently with performing a task can change behavior, invalidating conclusions about the original behavior. We minimized this issue by eliciting explanations just after each intervention was chosen, before its outcome was revealed. Additionally, we did not find any difference in the distribution of interventions on the free response trials and those chosen the first time participants identified the chains structure.

A second issue is that there are limits on the kinds of processes people can describe effectively in natural language, with rule based explanations being typically easier to express than those involving more complex statistical weighting and averaging. That is, even if someone weighed several factors in coming to a decision, they might explain this by mentioning only the most significant, or recently considered of these factors, falsely appearing to have relied on a one-reason decision strategy. There is an active debate about this, including suggestions that people's explanations for their choices are, in general, post-hoc rationalizations rather than genuine descriptions of process (Dennett, 1991; Johansson, Hall, Sikström, & Olsson, 2005), but also refutations of this interpretation (Newell & Shanks, 2014).

In sum, taken with appropriate caution, we suggest that this analysis does provides a valuable window on participants' subjective sense of their active testing, with their relatively specific focus on one aspect of the uncertainty at a time consistent with the idea that they rely on a mixture of heuristic questions.

The models pinned down interventions less tightly than beliefs in the sense that there was a great deal of spread in the individuals best fit across the models, and the proportional reductions in BIC were smaller. There are various possible reasons for this. Firstly, the models of belief change generally predicted one or few likely models, whereas there are typically many interventions of roughly equal informativeness to an ideal learner (see Figure 3), which could be performed in many different orders. This sets the bar for predictability for interventions much lower than for the causal judgments.

Secondly, to the extent that learners chose interventions based on a reduced encoding of the hypothesis space, we are also forced to average over our additional uncertainty about exactly which hypotheses or alternatives they were considering at the moment of choice (Markant & Gureckis, 2010).

A third issue is that of whether and how learners represented current uncertainty, and recruited this in choosing what to focus on. In the current work we assumed that learners were somewhat able to track the current local uncertainties and use these to choose what to target next. The modeling revealed that relative to the local intervention schema, the majority of participants did tend to focus on the areas of high current uncertainty (shown by the predominantly positive ρ in Figure 13 d) but we do not yet have a model for how they did this. It is plausible that learners used a heuristic to estimate their local confidence. For example, a sim-

ple option would be to accrue confidence in an edge, (or analogously in the the descendants of a variable or in the current hypothesis) for every search step for which it is considered and remains unchanged, reducing confidence every time it changes. In this way confidence in locales that survive more data and search become stronger, approximately mimicking reduction in local uncertainty.

We considered just three of a multitude of possible choices of local focus. These encompass most extant proposals for human search heuristics, encapsulating modular (Markant et al., 2015) constraint seeking (Ruggeri & Lombrozo, 2014) and confirmatory (Klayman & Ha, 1989) testing, placing all three within a unified schema and also showing that many learners dynamically switch between them.

Participants' free responses provided a complementary perspective, suggesting that even initial tests were generated as solutions to uncertainty about some specific subpart of the overall uncertainty space – often the descendants of some particular variable or the presence of some particular connection. This suggests that the the most important step in an intervention selection may not be the final choice of action but the prior choice of what to focus on next. This is captured in our model, under which the values of different interventions for a chosen focus do not depend on D^{t-1}. This means learners need not do extensive prospective calculation on every test but can learn gradually, for instance through experience and preplay (Pfeiffer & Foster, 2013), which interventions are likely to be informative relative to generic types of local focus. This knowledge could then be transferred to subsequent tests, and translated to tests with different targets – e.g. if $Do[x = 1]$ is effective for identifying the effects of x then $Do[y = 1]$ will be effective for identifying the effects of y.

It is worth noting from these data that even when participants' interventions were relatively uninformative from the perspective of ideal or even our heuristic learners, their explanations would generally reveal that they *were* informative with respect to some other question or source of uncertainty. For example, participants' tests that were uninformative with respect to identifying structure were often revealed, through our free response coding, to have been motivated by a desire to reduce uncertainty about internal w_S or background w_B noise.[20] From this perspective we might think of even the completely uninformative intervention choices (e.g. fixing all the variables) as legitimate tests of illegitimate hypotheses – e.g.hypotheses that were outside of the space of possibilities we intended participants to consider – such as whether fixed variables actually always took the states they were fixed to. More research is needed to explicate these internal steps leading up to an active learning action, but the implication based on the current research is that the solution will not require that the learner evaluate all possible outcomes of all possible actions under all possible models, but rather reflect a mixture of heuristics that can guide the gradual improvement of the learner's current theory.

[20]We might have extended the computational model of Bayesian inference to incorporate joint inference over models and parameters which would have incorporated this aspect of testing. However, this would have complicated analyses since participants were ultimately only incentivised to identify the right connections

The navy of one

At the start we argued that our Neurath's ship model could be seen as a single particle combined with an MCMC search. As such, we are claiming *Neurath's ship* as a form of boundedly rational approximate Bayesian inference. However, it is important to consider the point at which an approximation becomes so degenerate that it is merely a complicated way to describe a simple heuristic. Many would argue that this line is crossed long before reaching particle filters containing a single particle, or Markov chains lasting only 1 or 2 steps. It is certainly a leap to claim that such a process is calculating a proper posterior.

One alternative to starting from a normative computational level account and accepting a distant algorithmic approximation, is to start from the algorithm, i.e., the simple rules, and consider a computational account such as satisficing (Simon, 1982) that provides adequate license. Our account shares two important problems with this, but avoids two others.

One shared problem is the provenance of the rules - i.e., the situation-specific heuristics. We saw this in the manifold choice of local foci for the choice of intervention – we do not have an account of whence these hail. This is a common problem in the context of the adaptive toolbox (Gigerenzer, 2001) – it is hard to have a theory of the collection of tools.

A second shared problem follows on from this - namely how to choose which rule to apply under which circumstance. In our case, this is evident again in the mixtures of local focus rules – we were not able to provide a satisfying account of how participants make their selection of focus on a particular trial. The meta-problem of choosing the correct heuristic is again a common issue for satisficing approaches.

By contrast with a toolbox approach, though, our account smoothly captures varying degrees of sophistication between individuals. For instance, with the *Take the best* heuristic, Gigerenzer, Todd, and ABC Research Group (1999) give an attractive description of one-reason decision making that often outperforms regression in describing people's decisions from multiple cues. However, subsequent analyses have revealed that participants behave somewhere between the two (Newell & Shanks, 2003; Parpart, Jones, & Love, in revision) often using more than one cue, but certainly less than all the information available. Thus to understand their processing we must be able to express the halfway houses between ideal and overly simplistic processing (Lieder & Griffiths, 2015). In the same way, the approximate Bayesian perspective allows us to express different levels of approximation lying between fully probabilistic and fully heuristic processing, with the simplest form of Neurath's ship lying at the heuristic end of this road.

A further benefit of our account is the ease of generalization between tasks. Heuristic models are typically designed for, and are competent at, specific paradigms. Since they lack a more formal relationship with approximate rationality, they are hard to combine or often to apply in different or broader circumstances.

Here, we assumed that learners made updates at the level of individual directed edges. Again this is just one illustrative choice, but our model is consistent with the idea that the learners altered beliefs by making changes local to arbitrary sub-spaces of an unmanageable learning problem. We showed that so long as the learner's updates are conditioned on the

rest of their model, and are appropriately balanced, the connection to approximate Bayesian inference can be maintained through the ideas of MCMC sampling and a single particle particle filter. A sophisticated learner might be able to update several edges of their causal model at a single time, with a more complex proposal distribution. However, on a larger scale this is still likely to be a small subset of all potential relata that learner has encountered, meaning even the most sophisticated learner must lean on their broader beliefs for support.

In lower level cognition, inference takes place over simple quantities like magnitudes and is certainly probabilistic in the sense that humans can achieve near optimal integration of noisy signals in a variety of tasks including estimation (**?**) and motor control (e.g. **?**). At the top end of higher level cognition we have a global world-view, and explicit reasoning characterized by its single track nature. Rather than claiming these are completely different processes (Evans, 2003), the approximate probabilistic inference perspective can accommodate the whole continuum. At the lower level the brain can average over many values, as in particle filtering (Abbott & Griffiths, 2011), with a whole fleet of Neurath's ships, or via lots of long chains (Gershman et al., 2012; Lieder et al., 2012). In higher level cognition, however, the hypothesis space becomes increasingly unwieldy, and inference becomes increasingly approximate as it must rely on smaller fleets, i.e., fewer hypotheses, and more local alterations in the face of evidence. At the very top we have a navy of one, grappling with a single global model that can only be updated incrementally. It is worth noting that individuals can then play the role of particles again in group behavior (Courville & Daw, 2007), giving us approximate inference all the way up.

In sum, retaining the Bayesian machinery is valuable even as it becomes degenerate, because it allows us to express heuristic behavior without resorting to separate process model or abandoning close connections to an appropriate computational level understanding.

Scope of the theory

We modeled causal belief change as a process of gradually updating a single representation through local, conditional edits. While we chose to focus on causal structure inference within the causal Bayes net framework here, there is no reason why this approach should be limited to this domain. By taking the *Neurath's ship* metaphor to reveal an intuitive answer as to how people sidestep the intractability of rational theory formation (van Rooij et al., 2014), we can start to build more realistic models of how people generate the theories that they do and how and why they get stuck. We might explain the induction and adaptation of many of the rich representations utilized in cognition by analogous processes. Future work could explore the piecemeal induction of models involving multinomial, continuous (Nodelman, Shelton, & Koller, 2002; Pacer & Griffiths, 2011) or latent variables (Lucas, Holstein, & Kemp, 2014); unrestricted functional forms (Griffiths, Lucas, Williams, & Kalish, 2009); hierarchical organization (Griffiths & Tenenbaum, 2009; Williamson & Gabbay, 2005); and temporal (Pacer & Griffiths, 2012) and spatial (Battaglia, Hamrick, & Tenenbaum, 2013; Ullman et al., 2012; Ullman, Stuhlmüller, Goodman, & Tenenbaum, 2014) semantics. We are currently exploring the combination of production rules (Goodman, Tenenbaum, Feldman, & Griffiths, 2008) and local search to model discovery of new hypotheses in situations where the space of possibili-

ties is theoretically infinite. The sequential conditional re-evaluation process illustrated by our *Neurath's ship* model shows how this radical antifoundationalism need not be fatal for theory building in general.

Conclusions

In this paper, we proposed a new model of causal theory change, based on an old idea from philosophy of science – that learners cannot maintain a distribution over all possible beliefs, and so must rely on sequential local changes to a single representation when updating beliefs to incorporate new evidence. We showed that we can provide a good account of participants' sequences of judgments in three experiments and argued that our model offers a flexible candidate for explaining how complex representations can be formed in cognition. We also analyzed participants' information-gathering behavior, finding it consistent with the thesis that learners focus on resolving manageable areas of local uncertainty rather than global uncertainty, showing cognizance of their learning limitations. Together these accounts show how people manage to construct rich, causally-structured representations through their interactions with a complex noisy world.

Acknowledgments

We would like to thank Falk Lieder, Michael Pacer, Paula Parpart, Eric Schulz and Maarten Speekenbrink and our anonymous reviewers for many helpful comments and suggestions. NB is supported by a 4-year Engineering and Physical Sciences Research Council UK stipend awarded by UCL's Centre for Doctoral Training in Financial Computing and Analytics. DL is supported by an Economic and Social Research Council UK grant (RES 062330004). TG is supported by the Air Force Office of Scientific Research (FA9550-13-1-0170). PD is supported by the Gatsby Charitable Foundation.

References

Abbott, J. T., Austerweil, J. L., & Griffiths, T. L. (2012). Human memory search as a random walk in a semantic network. In *Nips* (pp. 3050–3058).

Abbott, J. T., & Griffiths, T. L. (2011). Exploring the influence of particle filter parameters on order effects in causal learning. In *Proceedings of the 33rd annual conference of the cognitive science society.*

Anderson, J. (1990). *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Austerweil, J. L., & Griffiths, T. L. (2011). Seeking confirmation is rational for deterministic hypotheses. *Cognitive Science*, *35*(3), 499–526.

Battaglia, P. W., Hamrick, J. B., & Tenenbaum, J. B. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences*, *110*(45), 18327–18332.

Bishop, C. M. (2006). *Pattern recognition and machine learning* (Vol. 4) (No. 4). Springer, New York.

Bonawitz, E., Denison, S., Gopnik, A., & Griffiths, T. L. (2014). Win-stay, lose-sample: A simple sequential algorithm for approximating bayesian inference. *Cognitive psychology*, *74*, 35–65.

Bramley, N. R., Dayan, P., & Lagnado, D. A. (2015). Staying afloat on Neurath's boat–Heuristics for sequential causal learning. In *Proceedings of the 37th annual conference of the cognitive science society* (pp. 262–267). Austin, TX: Cognitive Science Society.

Bramley, N. R., Gerstenberg, T., & Lagnado, D. A. (2014). The order of things: Inferring causal structure from temporal patterns. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th annual conference of the cognitive science society* (p. 236-242). Austin, TX: Cognitive Science Society.

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

Bramley, N. R., Nelson, J. D., Speekenbrink, M., & Lagnado, D. A. (2014). *What should causal learners value.* Poster presented at the Psychonomic Society Annual Meeting 2014. (to appear)

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

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's mechanical turk a new

source of inexpensive, yet high-quality, data? *Perspectives on psychological science*, *6*(1), 3–5.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, *104*, 367-405.

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

Cooper, G. F. (1990). The computational complexity of probabilistic inference using bayesian belief networks. *Artificial intelligence*, *42*(2), 393–405.

Cooper, G. F., & Herskovits, E. (1992). A Bayesian method for the induction of probabilistic networks from data. *Machine learning*, *9*(4), 309–347.

Courville, A. C., & Daw, N. D. (2007). The rat as particle filter. In *Advances in neural information processing systems* (pp. 369–376).

Crump, M. J., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating amazon's mechanical turk as a tool for experimental behavioral research. *PloS one*, *8*(3), e57410.

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

DeCarlo, L. T. (1992). Intertrial interval and sequential effects in magnitude scaling. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(4), 1080.

Dennett, D. C. (1991). *Consciousness explained*. London, UK: Penguin.

Dobson, A. J. (2010). *An introduction to generalized linear models*. CRC press.

Dolan, R. J., & Dayan, P. (2013). Goals and habits in the brain. *Neuron*, *80*(2), 312–325.

Duhem, P. M. M. (1991). *The aim and structure of physical theory* (Vol. 13). Princeton University Press.

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

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological review*, *87*(3), 215.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis.* Cambridge, MA: MIT Press.

Evans, J. S. B. (2003). In two minds: dual-process accounts of reasoning. *Trends in cognitive sciences*, *7*(10), 454–459.

Fernbach, P. M., & Sloman, S. A. (2009). Causal learning with local computations. *Journal of experimental psychology: Learning, memory, and cognition*, *35*(3), 678.

Geman, S., & Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*(6), 721–741.

Georges, A., Kotliar, G., Krauth, W., & Rozenberg, M. J. (1996). Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimen-

sions. *Reviews of Modern Physics*, *68*(1), 13.

Gershman, S. J., Vul, E., & Tenenbaum, J. B. (2012). Multistability and perceptual inference. *Neural computation*, *24*(1), 1–24.

Gigerenzer, G. (2001). The adaptive toolbox. In G. Gigerenzer & R. Selten (Eds.), *Bounded rationality* (pp. 37–50). USA: MIT Press.

Gigerenzer, G., Todd, P. M., & ABC Research Group. (1999). *Simple heuristics that make us smart*. Oxford University Press New York.

Gilden, D. L. (2001). Cognitive emissions of 1/f noise. *Psychological review*, *108*(1), 33.

Goodman, N. D., Tenenbaum, J. B., Feldman, J., & Griffiths, T. L. (2008). A rational analysis of rule-based concept learning. *Cognitive Science*, *32*(1), 108–154.

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? a comparative analysis of six preconceptions about internet questionnaires. *American Psychologist*, *59*(2), 93.

Goudie, R. J., & Mukherjee, S. (2011). *An efficient gibbs sampler for structural inference in bayesian networks* (Tech. Rep.). Citeseer.

Griffiths, T. L., Lucas, C., Williams, J., & Kalish, M. L. (2009). Modeling human function learning with gaussian processes. In *Advances in neural information processing systems* (pp. 553–560).

Griffiths, T. L., & Tenenbaum, J. (2007). Two proposals for causal grammars. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy and computation.* Oxford: Oxford University Press.

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

Gureckis, T. M., & Markant, D. B. (2009). Active learning strategies in a spatial concept learning game. *... of the 31st Annual Conference of the ...*. Retrieved from `http://gureckislab.org/papers/GureckisMarkantCogSci2009.pdf`

Hamrick, J. B., Smith, K. A., Griffiths, T. L., & Vul, E. (2015). Think again? the amount of mental simulation tracks uncertainty in the outcome. In *Proceedings of the thirty-seventh annual conference of the cognitive science society.*

Hauser, D. J., & Schwarz, N. (2015). Attentive turkers: Mturk participants perform better on online attention checks than do subject pool participants. *Behavior research methods*, 1–8.

Holyoak, K. J., & Cheng, P. W. (2011, January). Causal learning and inference as a rational process: the new synthesis. *Annual review of psychology*, *62*, 135–63.

Hwang, C.-R. (1988). Simulated annealing: theory and applications. *Acta Applicandae Mathematicae*, *12*(1), 108–111.

Jaakkola, T. S. (2001). 10 tutorial on variational approximation methods. *Advanced*

mean field methods: theory and practice, 129.

James, W. (2004). *The principles of psychology (volume 1 of 2)*. Digireads. com Publishing.

Johansson, P., Hall, L., Sikström, S., & Olsson, A. (2005). Failure to detect mismatches between intention and outcome in a simple decision task. *Science*, *310*(5745), 116–119.

Julier, S. J., & Uhlmann, J. K. (1997). New extension of the kalman filter to nonlinear systems. In *Aerosense'97* (pp. 182–193).

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.

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

Klayman, J., & Ha, Y.-w. (1989). Hypothesis testing in rule discovery: Strategy, structure, and content. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(4), 596.

Krippendorff, K. (2012). *Content analysis: An introduction to its methodology*. Sage.

Kushnir, T., Gopnik, A., Lucas, C., & Schulz, L. (2010, January). Inferring hidden causal structure. *Cognitive science*, *34*(1), 148–60.

Lagnado, D. A., & Sloman, S. (2002). Learning causal structure. In W. Gray & C. D. Schunn (Eds.), *Proceedings of the twenty-fourth annual conference of the cognitive science society, mahwah,*. NJ: Erlbaum.

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

Lagnado, D. A., & Sloman, S. A. (2006). Time as a guide to cause. *Journal of experimental psychology. Learning, memory, and cognition*, *32*(3), 451–60.

Lagnado, D. A., Waldmann, M., Hagmayer, Y., & Sloman, S. (2007). Beyond covariation: cues to causal structure. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (p. 154-72). London: Oxford University Press.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, 159–174.

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

Lieder, F., Griffiths, T., & Goodman, N. (2012). Burn-in, bias, and the rationality of anchoring. In *Advances in neural information processing systems* (pp. 2690–2798).

Lieder, F., & Griffiths, T. L. (2015). When to use which heuristic: A rational solution

to the strategy selection problem. In *Proceedings of the 37th annual conference of the cognitive science society* (pp. 1362–1367).

Lieder, F., Griffiths, T. L., Huys, Q. J. M., & Goodman, N. D. (under review). The anchoring bias reflects rational use of cognitive resources. *Psychological Review*.

Liu, J. S., & Chen, R. (1998). Sequential monte carlo methods for dynamic systems. *Journal of the American statistical association*, *93*(443), 1032–1044.

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

Lucas, C. G., Bridgers, S., Griffiths, T. L., & Gopnik, A. (2014). When children are better (or at least more open-minded) learners than adults: Developmental differences in learning the forms of causal relationships. *Cognition*, *131*(2), 284–299.

Lucas, C. G., Holstein, K., & Kemp, C. (2014). Discovering hidden causes using statistical evidence..

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

Madigan, D., & Raftery, A. E. (1994). Model selection and accounting for model uncertainty in graphical models using occam's window. *Journal of the American Statistical Association*, *89*(428), 1535–1546.

Madigan, D., York, J., & Allard, D. (1995). Bayesian graphical models for discrete data. *International Statistical Review/Revue Internationale de Statistique*, 215–232.

Markant, D. B., & Gureckis, T. M. (2010). Category learning through active sampling. *Proceedings of the of the 32nd Annual Meeting of the Cognitive Science Society*, 248–253.

Markant, D. B., & Gureckis, T. M. (2012). Does the utility of information influence sampling behavior? *Proceedings of the 34th Annual Conference of ...*. Retrieved from `http://nyuccl.org/papers/MarkantGureckis.CogSci2012.battleship.pdf`

Markant, D. B., Settles, B., & Gureckis, T. M. (2015). Self-directed learning favors local, rather than global, uncertainty. *Cognitive science*.

Marr, D. (1982). *Vision*. New York: Freeman & Co.

Mason, W., & Suri, S. (2012). Conducting behavioral research on amazon's mechanical turk. *Behavior research methods*, *44*(1), 1–23.

McCormack, T., Bramley, N. R., Frosch, C., Patrick, F., & Lagnado, D. A. (2016). Children's use of interventions to learn causal structure. *Journal of Experimental Child Psychology*, *141*, 1-22.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, *21*(6), 1087–1092.

Murphy, K. P. (2001). *Active learning of causal Bayes net structure* (Tech. Rep.). UC Berkeley.

Navarro, D. J., & Perfors, A. F. (2011). Hypothesis generation, sparse categories, and the positive test strategy. *Psychological review*, *118*(1), 120.

Nelson, J. D. (2005, October). Finding useful questions: on Bayesian diagnosticity, probability, impact, and information gain. *Psychological review*, *112*(4), 979–99. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16262476 doi: 10.1037/0033-295X.112.4.979

Nelson, J. D., Divjak, B., Gudmundsdottir, G., Martignon, L. F., & Meder, B. (2014). Children?s sequential information search is sensitive to environmental probabilities. *Cognition*, *130*(1), 74–80.

Neurath, O. (1932/1983). Protocol statements. In *Philosophical papers 1913–1946* (pp. 91–99). Springer.

Newell, B. R., & Shanks, D. R. (2003). Take the best or look at the rest? Factors influencing "one-reason" decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(1), 53–65. doi: 10.1037/0278-7393.29.1.53

Newell, B. R., & Shanks, D. R. (2014). Unconscious influences on decision making: A critical review. *Behavioral and Brain Sciences*, *37*(01), 1–19.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology*, *2*(2), 175.

Nielsen, F., & Nock, R. (2011). A closed-form expression for the sharma–mittal entropy of exponential families. *Journal of Physics A: Mathematical and Theoretical*, *45*(3), 032003.

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

Nodelman, U., Shelton, C. R., & Koller, D. (2002). Continuous time bayesian networks. In *Proceedings of the eighteenth conference on uncertainty in artificial intelligence* (pp. 378–387).

Nyberg, E., & Korb, K. (2006). Informative interventions. In F. Russo & J. Williamson (Eds.), *Causality and probability in the sciences.* College Publications: London.

Pacer, M., & Griffiths, L. (2012). Elements of a rational framework for continuous-time causal induction. In *Proc. annu. conf. cogn. sci. soc* (Vol. 1, pp. 833–838).

Pacer, M., & Griffiths, T. L. (2011). A rational model of causal induction with continuous causes. In *Advances in neural information processing systems* (pp. 2384–2392).

Parpart, P., Jones, M., & Love, B. (in revision). Heuristics as bayesian inference.

Patil, K. R., Zhu, X., Kopeć, Ł., & Love, B. C. (2014). Optimal teaching for limited-capacity human learners. In *Advances in neural information processing systems* (pp.

2465–2473).

Pearl, J. (2000). *Causality*. New York: Cambridge University Press (2nd edition).

Petrov, A. A., & Anderson, J. R. (2005). The dynamics of scaling: a memory-based anchor model of category rating and absolute identification. *Psychological review*, *112*(2), 383.

Pfeiffer, B. E., & Foster, D. J. (2013). Hippocampal place-cell sequences depict future paths to remembered goals. *Nature*, *497*(7447), 74–79.

Quine, W. v. O. (1969). *Word and object*. MIT press.

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

Rescorla, R. A., Wagner, A. R., et al. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical conditioning II: Current research and theory*, *2*, 64–99.

Robinson, R. W. (1977). Counting unlabeled acyclic digraphs. In *Combinatorial mathematics v* (pp. 28–43). Springer.

Ruggeri, A., & Lombrozo, T. (2014). Learning by asking: How children ask questions to achieve efficient search. In *36th annual meeting of the cognitive science society* (pp. 1335–1340).

Russo, J. E., Johnson, E. J., & Stephens, D. L. (1989). The validity of verbal protocols. *Memory & cognition*, *17*(6), 759–769.

Sanborn, A. N. (2015). Types of approximation for probabilistic cognition: Sampling and variational. *Brain and cognition*.

Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2010). Rational approximations to rational models: alternative algorithms for category learning. *Psychological review*, *117*(4), 1144.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, *6*(2), 461–464.

Settles, B. (2012). Active learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, *6*(1), 1–114.

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

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

Sobel, D. M., & Kushnir, T. (2006). The importance of decision making in causal learning from interventions. *Memory & Cognition*, *34*(2), 411–419.

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

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

Treisman, M., & Williams, T. C. (1984). A theory of criterion setting with an application to sequential dependencies. *Psychological Review*, *91*(1), 68.

Tsamardinos, I., Brown, L. E., & Aliferis, C. F. (2006). The max-min hill-climbing bayesian network structure learning algorithm. *Machine learning*, *65*(1), 31–78.

Ullman, T. D., Goodman, N. D., & Tenenbaum, J. B. (2012). Theory learning as stochastic search in the language of thought. *Cognitive Development*, *27*(4), 455–480.

Ullman, T. D., Stuhlmüller, A., Goodman, N., & Tenenbaum, J. (2014). Learning physics from dynamical scenes. In *Proceedings of the thirty-sixth annual conference of the cognitive science society.*

van Rooij, I., Wright, C. D., Kwisthout, J., & Wareham, T. (2014). Rational analysis, intractability, and the prospects of 'as if'-explanations. *Synthese*, 1–20.

Vul, E., Goodman, N. D., Griffiths, T. L., & Tenenbaum, J. B. (2009). One and done? optimal decisions from very few samples. In *Proceedings of the 31st annual conference of the cognitive science society* (Vol. 1, pp. 66–72).

Waldmann, M., Cheng, P., Hagmayer, Y., & Blaisdell, A. (2008). Causal learning in rats and humans: A minimal rational model. In M. Oaksford & N. Chater (Eds.), *The probabilistic mind.* Oxford: OUP.

Weierstrass, K. (1902). *Mathematische werke*.

Williamson, J., & Gabbay, D. (2005). Recursive causality in bayesian networks and self-fibring networks. *Laws and models in the sciences*, 173–221.

Yeung, S., & Griffiths, T. L. (2011). Estimating human priors on causal strength. In *Proceedings of the 33rd annual conference of the cognitive science society* (pp. 1709–1714).

Yu, A., & Dayan, P. (2003). Expected and unexpected uncertainty: ACh and NE in the neocortex. *Advances in neural information processing systems*, 173–180.

A Formal specification of the models

Representation and inference

A noisy-OR parametrized causal model m over variables X, with strength and background parameters w_S and w_B assign a likelihood to each datum (a complete observation, or the outcome of an intervention) $\mathbf{d}$ as the product of the probability of each variable that was not intervened upon given the states of its parents

$$P(\mathbf{d}|m, \mathbf{w}) = \prod_{x \in X} P(x|\mathbf{d}_{pa(x)}, \mathbf{w}) \tag{22}$$

$$P(x|\mathbf{d}_{pa(x)}, \mathbf{w}) = x + (1 - 2x)(1 - w_B)(1 - w_S)^{\sum_{y \in pa(x)} y} \tag{23}$$

where $pa(x)$ denotes the parents of variable x in the causal model (see Figure 1 for an example). We can thus compute the posterior probability of model $m \in M$ over a set of models $\mathcal{M}$ given a prior $P(M)$ and data $D = \{\mathbf{d}^i\}$ associated with interventions $C = \{\mathbf{c}^i\}$. We can condition on w_S and w_B if known (e.g. in Experiment 1)

$$P(m|D, \mathbf{w}) = \frac{P(D|m, \mathbf{w}; C)P(m)}{\sum_{m' \in M} P(D|m', \mathbf{w}; C)P(m')} \tag{24}$$

or else marginalize over their possible values (e.g. in Experiment 2)

$$P(m|D) = \frac{\int_{\mathbf{w}} P(D|m, \mathbf{w}; C)p(\mathbf{w})P(m)\, \mathrm{d}\mathbf{w}}{\sum_{m' \in M} \int_{\mathbf{w}} P(D|m', \mathbf{w}; C)p(\mathbf{w})P(m')\, \mathrm{d}\mathbf{w}} \tag{25}$$

Intervention choice

The value of an intervention can be quantified relative to a notion of uncertainty. Here we adopt Shannon entropy (**?**), for which the uncertainty in a distribution over causal models M is given by

$$H(M) = -\sum_{m \in M} P(m) \log_2 P(m) \tag{26}$$

Assuming $\mathbf{w}$ is known, let $\Delta H(M|\mathbf{d}, \mathbf{w}; \mathbf{c})$ refer to the reduction in uncertainty going from prior $P(M)$ to posterior $P(M|\mathbf{d}, \mathbf{w}; \mathbf{c})$ after performing intervention c, then seeing data $\mathbf{d}$

$$\Delta H(M|d, \mathbf{w}; \mathbf{c}) = \left[-\sum_{m \in M} P(m) \log P(m)\right] - \left[-\sum_{m \in M} P(m|\mathbf{d}, \mathbf{w}; \mathbf{c}) \log P(m|\mathbf{d}, \mathbf{w}; \mathbf{c})\right] \tag{27}$$

Given this objective, we can define the value of an intervention as the expected reduction in uncertainty after seeing its outcome. To get the expectancy, we must average, prospectively, over the different possible outcomes $\mathbf{d} \in \mathcal{D}_{\mathbf{c}}$ (where $\mathcal{D}_{\mathbf{c}}$ is the space of possible outcomes of

intervention $\mathbf{c}$) weighted by their marginal likelihoods under the prior, giving

$$\mathop{\mathbb{E}}_{\mathbf{d}\in\mathcal{D}_\mathbf{c}}[\Delta H(M|\mathbf{d},\mathbf{w};\mathbf{c})] = \sum_{\mathbf{d}\in\mathcal{D}_\mathbf{c}}\left(\Delta H(M|\mathbf{d},\mathbf{w};\mathbf{c})\sum_{m\in M}P(\mathbf{d}|m,\mathbf{w};\mathbf{c})P(m)\right) \tag{28}$$

For a greedily optimal sequence of interventions $\mathbf{c}^1,\ldots,\mathbf{c}^t$, we take $P(M|D^{t-1},\mathbf{w};C^{t-1})$ as $P(M)$ and $P(M|D^t,\mathbf{w};C^{t-1},\mathbf{c}^t)$ as $P(M|\mathbf{d},\mathbf{w};\mathbf{c})$ in Equation 27. The most valuable intervention at a given time point is then

$$\mathbf{c}^t = \arg\max_{\mathbf{c}\in\mathcal{C}}\mathop{\mathbb{E}}_{\mathbf{d}\in\mathcal{D}_\mathbf{c}}\left[\Delta H(M|\mathbf{d},D^{t-1},\mathbf{w};C^{t-1},\mathbf{c})\right] \tag{29}$$

If $\mathbf{w}$ is unknown, we must use the marginal distribution, replacing Equation 27 with

$$\Delta H(M|d;\mathbf{c}) = \left[-\sum_{m\in M}P(m)\log P(m)\right] - \left[-\sum_{m\in M}\int_\mathbf{w}P(m|\mathbf{d},\mathbf{w};\mathbf{c})p(\mathbf{w})\,\mathrm{d}\mathbf{w}\log\int_\mathbf{w}P(m|\mathbf{d},\mathbf{w};\mathbf{c})p(\mathbf{w})\,\mathrm{d}\mathbf{w}\right] \tag{30}$$

An algorithmic-level model of sequential belief change

Let E be an adjacency matrix such that the upper triangle entries where E_{ij} (if $i<j\leq N$) denotes the state of edge $i-j$ in a causal model m. Any model $m\in M$ corresponds to a setting for all E_{ij} where $i<j\leq N$, to one of three edge states $e\in\{1:i\rightarrow j,\ 0:i\nleftrightarrow j,\ -1:i\leftarrow j\}$. By starting with any hypothesis and iteratively sampling from the conditional distributions on edge states $P(E_{ij}|E_{\setminus ij},\mathcal{D}_r^t,\mathbf{w};\mathcal{C}_r^t)$ (Goudie & Mukherjee, 2011) using the following equation:

$$P(E_{ij}=e|E_{\setminus ij},\mathcal{D}_r^t,\mathbf{w};\mathcal{C}_r^t) = \frac{P(E_{ij}=e|E_{\setminus ij},\mathcal{D}_r^t,\mathbf{w};\mathcal{C}_r^t)}{\sum_{e'\in E_{ij}}P(E_{ij}=e'|E_{\setminus ij},\mathcal{D}_r^t,\mathbf{w};\mathcal{C}_r^t)} \tag{31}$$

we can cheaply generate chains of dependent samples from $P(M|\mathcal{D}_r^t,\mathbf{w};\mathcal{C}_r^t)$. This can be done systematically (cycling through all edges $\in i<j\leq N$), or randomly selecting the next edge sample with $P(\frac{1}{|i,j|})$ where $|i,j|$ is the number of edges in the graph. Here we assume random sampling for simplicity. Thus, on each step, the selected E_{ij} is updated using the newest values of $E_{\setminus ij}$.[21] Specifically, we assume that after each new piece of evidence arrives:

1. The learner begins sampling with edges $E_{ij}^{(0)}$ for all i and j set as they were in their previous judgment b^{t-1}.
2. They then randomly select an edge E_{ij} in $i<j\leq N$ to update.
3. They resample $E_{ij}^{(1)}$ using Equation 31.
4. If the search does not result in a new model they keep collecting evidence $\mathcal{D}_r^t=\{\mathcal{D}_r^t,\mathbf{d}^t\}$, $\mathbf{c}=\{\mathcal{C}_r^t,\mathbf{c}^t\}$. If it does, the evidence is used up and forgotten, and they begin collecting evidence again (e.g. resetting $\mathcal{D}_r^t=\{\}$ and $\mathcal{C}_r^t=\{\}$).
5. The learner repeats steps 1 to 4 k times, with their final edge choices $E^{(k)}$ constituting their new belief b^t.

[21] Edge changes that would create a cyclic graph always have a probability of zero

We assume for simplicity that b^0, before any data has been seen is an unconnected graph, but have tested this assumption by fitting the data from t=2 onward only finding better fits overall and a stronger win for *Neurath's ship* over the other models we consider.

Resampling, hill climbing or random change

We also consider generalizations of Equation 31 allowing transitions to be governed by higher powers of $P(E_{ij} = e|E_{\setminus ij}, \mathcal{D}_r^t, \mathbf{w}; \mathcal{C}_r^t)$

$$P^\omega(E_{ij} = e|E_{\setminus ij}, \mathcal{D}_r^t, \mathbf{w}; \mathcal{C}_r^t) = \frac{P^\omega(E_{ij} = e|E_{\setminus ij}, \mathcal{D}_r^t, \mathbf{w}; \mathcal{C}_r^t)}{\sum_{e' \in E_{ij}} P^\omega(E_{ij} = e'|E_{\setminus ij}, \mathcal{D}_r^t, \mathbf{w}; \mathcal{C}_r^t)} \tag{32}$$

yielding stronger preference for the most likely state of e_{ij} if $\omega > 1$ and more random sampling if $\omega < 1$.

A distribution over search lengths

We assume that for each update, the learner's length of search k is drawn from a Poisson distribution with average $\lambda \in [0, \infty]$

$$P(k) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{33}$$

Putting these together

To calculate the probability distribution of new belief b^t given $\mathbf{d}^t$, b^{t-1} search behavior ω and a chain of length k, we first construct the transition matrix R_t^ω for the Markov search chain by averaging over the conditional distributions associated with the choice of each edge, weighted by the probability of selecting that edge

$$R_t^\omega = \sum_{i<j \leq N} P^\omega(E_{ij} = e|E_{\setminus ij}, \mathcal{D}_r^t, \mathbf{w}; \mathcal{C}_r^t) \times \frac{1}{|i, j|} \tag{34}$$

for each possible belief b.

By raising this transition matrix to the power k (i.e. some search length) and selecting the row corresponding to starting belief $[(R_t^\omega)^k]_{b^{t-1}}$, we get the probability of adopting each $m \in M$ as new belief b^t (see Figure 2 for a visualization) at the end of the k length search

$$P(B^t|\mathcal{D}_r^t, b^{t-1}, \omega, k; \mathcal{C}_r^t) = [(R_t^\omega)^k]_{b^{t-1}m} \tag{35}$$

Finally, by averaging over different possible chain lengths k, weighted by their probability Poisson(λ) we get the marginal probability that a learner will move to each possible new belief in B at t

$$P(B^t|\mathcal{D}_r^t, b^{t-1}, \omega, \lambda; \mathcal{C}_r^t) = \sum_0^\infty \frac{\lambda^k e^{-\lambda}}{k!} [(R_t^\omega)^k]_{b^{t-1}m} \tag{36}$$

A local uncertainty schema

Edge focus

Relative to a focus on an edge E_{xy}, intervention values were calculated using expected information as in Appendix A, but assuming prior entropy as that of a uniform distribution over the three possible edge states

$$H(E_{xy}|E_{\setminus xy}) = -3\left(\frac{1}{3}\log_2\frac{1}{3}\right) \tag{37}$$

and calculating posterior entropies for the possible outcomes $\mathbf{d} \in \mathcal{D}$ using

$$H(E_{xy}|E_{\setminus xy}, \mathbf{d}, \mathbf{w}; \mathbf{c}) = -\sum_{z\in\{-1,0,1\}} P(E_{xy} = z|E_{\setminus xy}, \mathbf{d}, \mathbf{w}; \mathbf{c}) \log_2 P(E_{xy} = z|E_{\setminus xy}, \mathbf{d}, \mathbf{w}; \mathbf{c}) \tag{38}$$

Effect focus entropy

Relative to a focus on the effects of variable x, intervention values were calculated using expected information as in Appendix A but using prior entropy, calculated by partitioning a uniform prior over models M into sets of models $\mathrm{Mo}(z)$ corresponding to each descendant set $z \subseteq \mathrm{De}(x)$

$$H(\mathrm{De}(x)) = -\sum_{z\subseteq\mathrm{De}(x)} \left(\sum_{m\in\mathrm{Mo}(z)} \frac{1}{|M|}\right) \log_2 \left(\sum_{m\in\mathrm{Mo}(z)} \frac{1}{|M|}\right) \tag{39}$$

Posterior entropies were then calculated by summing over probabilities of the the elements in each $\mathrm{Mo}(z)$ for each $z \subseteq \mathrm{De}(x)$

$$H(\mathrm{De}(x)|\mathbf{d}, \mathbf{w}; \mathbf{c}) = -\sum_{z\subseteq\mathrm{De}(x)} \left(\sum_{m\in\mathrm{Mo}(z)} P(m|\mathbf{d}, \mathbf{w}; \mathbf{c})\right) \log_2 \left(\sum_{m\in\mathrm{Mo}(z)} P(m|\mathbf{d}, \mathbf{w}; \mathbf{c})\right) \tag{40}$$

Confirmation focus entropy

Relative to a focus on distinguishing current hypothesis b^t from null hypothesis b^0, intervention values were calculated using expected information as above but prior entropy was always based on a uniform prior over the two hypotheses

$$H(\{b^t, b^0\}) = -2\left(\frac{1}{2}\log_2\frac{1}{2}\right) \tag{41}$$

and posterior entropies were calculated using

$$H(\{b^t, b^0\}|\mathbf{d}, \mathbf{w}; \mathbf{c}) = -\sum_{z\in\{0,t\}} \frac{P(b^z|\mathbf{d}, \mathbf{w}; \mathbf{c})}{\sum_{z'\in\{0,t\}} P(b^{z'}|\mathbf{d}, \mathbf{w}; \mathbf{c})} \log_2 \frac{P(b^z|\mathbf{d}, \mathbf{w}; \mathbf{c})}{\sum_{z'\in\{0,t\}} P(b^{z'}|\mathbf{d}, \mathbf{w}; \mathbf{c})} \tag{42}$$

B Additional modeling details

All models were fit using maximum likelihood. Maximum likelihood estimates were found using Brent (for one parameter) or Nelder-Mead (for several parameters) optimization, as implemented by R's `optim` function. Convergence to global optima was checked by repeating all optimizations with a range of randomly selected starting parameters.

k For averaging across different values of k in the belief models, we capped k at 50 and renormalized the distribution such that $P(k \geq 0 \wedge k \leq 50) = 1$. This made negligible difference to the fits since the probabilities of $P(B^t|\mathbf{d}^r, b^{t-1}, \omega, k; \mathcal{C}_r^t)$ for values of $k \gg N$ (where N is the number of variables) were very similar.

ϵ To allow that participants are liable to occasionally lapse concentration or forget the outcome of a test, we included a lapse parameter ϵ – i.e., a parametric amount of decision noise $\epsilon \in [0, 1]$ – so that the probability of a belief would be a mixture of that predicted by the model and uniform noise. This ensured that occasional random judgments did not have undue effects on the other parameters of each model.

b^0 We assume for simplicity that people's starting belief, b^0, before any data has been seen, is an unconnected graph.

Marginalization

For all modeling in Experiment 3, we had to average over the unknown noise $\mathbf{w}$. To do this, we drew 1000 paired uniformly distributed w_S and w_B samples and averaged over these when computing marginal likelihoods and posteriors. These marginal priors and posteriors were used for computing expected information gain values.

Evaluating fits

Baseline acts as the null model for computing BIC's (Schwarz, 1978) and pseudo-R^2's (Dobson, 2010) for all other models.

Model recovery

Table 5: Belief Model Recovery Analysis

Bramley, Lagnado, and Speekenbrink (2015)													
All simulated participants							*Best fit participants*						
	Baseline	Rational	WSLS	SE	NS-RE	NS		Baseline	Rational	WSLS	SE	NS-RE	NS
Baseline	**134**	0	4	0	1	0	Random	**1**	0	0	0	0	0
Rational	39	**95**	4	0	0	1	Rational	0	**10**	0	0	0	0
WSLS	5	0	**133**	1	0	0	WSLS	0	0	**56**	0	0	0
SE	1	0	2	**119**	17	0	SE	0	0	0	**13**	2	0
NS-RE	2	0	1	53	**82**	1	NS-RE	0	0	1	9	**20**	0
NS	3	0	3	11	**72**	50	NS	0	0	1	1	**16**	9
Exp 1: Learning larger models													
All simulated participants							*Best fit participants*						
	Baseline	Rational	WSLS	SE	NS-RE	NS		Baseline	Rational	WSLS	SE	NS-RE	NS
Baseline	**114**	1	3	0	2	0	Baseline	**1**	0	0	0	0	0
Rational	42	**75**	2	0	1	0	Rational	0	0	0	0	0	0
WSLS	4	1	**115**	0	0	0	WSLS	1	1	**27**	0	0	0
SE	6	0	9	**94**	11	0	SE	0	0	1	**31**	0	0
NS-RE	2	0	1	3	**112**	2	NS-RE	0	0	0	1	**29**	1
NS	3	0	3	2	30	**82**	NS	0	0	0	0	4	**23**
Exp 2: Unknown strengths													
All simulated participants							*Best fit participants*						
	Baseline	Rational	WSLS	SE	NS-RE	NS		Baseline	Rational	WSLS	SE	NS-RE	NS
Baseline	**107**	1	3	0	0	0	Baseline	**2**	0	0	0	0	0
Rational	17	**92**	2	0	0	0	Rational	1	**18**	0	0	0	0
WSLS	7	0	**100**	2	2	0	WSLS	1	0	**16**	0	1	0
SE	9	0	6	**93**	3	0	SE	0	0	0	**30**	0	0
NS-RE	9	0	8	3	**89**	2	NS-RE	0	0	0	0	**5**	0
NS	5	2	9	2	**70**	23	NS	0	2	3	0	**22**	10

Note: Rows denote simulation rule, and columns the model used to fit the simulated choices. The number in each cell shows how many of the simulations using this rule were best fit by that model. Right hand side restricts this to simulations using the parameters taken from participants who were actually best fit by each model.