Journal of Experimental Psychology:
Learning, Memory, and Cognition
2015, Vol. 41, No. 3, 708–731

0278-7393/15/$12.00 http://dx.doi.org/10.1037/xlm0000061

Conservative Forgetful Scholars: How People Learn Causal Structure Through Sequences of Interventions

Neil R. Bramley, David A. Lagnado, and Maarten Speekenbrink
University College London

Interacting with a system is key to uncovering its causal structure. A computational framework for interventional causal learning has been developed over the last decade, but how real causal learners might achieve or approximate the computations entailed by this framework is still poorly understood. Here we describe an interactive computer task in which participants were incentivized to learn the structure of probabilistic causal systems through free selection of multiple interventions. We develop models of participants' intervention choices and online structure judgments, using expected utility gain, probability gain, and information gain and introducing plausible memory and processing constraints. We find that successful participants are best described by a model that acts to maximize information (rather than expected score or probability of being correct); that forgets much of the evidence received in earlier trials; but that mitigates this by being conservative, preferring structures consistent with earlier stated beliefs. We explore 2 heuristics that partly explain how participants might be approximating these models without explicitly representing or updating a hypothesis space.

Keywords: causal, learning, Bayesian, active, structure

By representing causal relationships, people are able to predict, control, and reason about their environment (Pearl, 2000; Sloman, 2005; Tenenbaum, Kemp, Griffiths, & Goodman, 2011). A large body of work in psychology has investigated how and when people infer these causal relationships from data (Cheng, 1997; Griffiths & Tenenbaum, 2009; Shanks, 1995; Waldmann, 2000; Waldmann & Holyoak, 1992). However, people are not passive data crunchers; they are active embodied agents who continually interact with and manipulate their proximal environment and thus partially control their data stream. This ability to self-direct and intervene on the environment makes it possible for people to efficiently bootstrap their own learning. This is known as *active learning.*

The idea that self-direction is crucial to learning is well established in education and frequently noted in developmental psychology, where self-directed "play" is seen as vital to healthy development (e.g., Bruner, Jolly, & Sylva, 1976; Piaget, 1930/2001). Additionally, recent work in cognitive science has shown that people learn categories and spatial concepts more quickly by actively selecting informative samples (e.g., Gureckis & Markant, 2009, 2012; Markant & Gureckis, 2010). However, it is in identifying causal structure that taking control of one's data stream becomes really essential (Pearl, 2000; Woodward, 2003). As an example, imagine you are interested in learning about the relationship between two variables *A* and *B.* Concretely, suppose you are a medical researcher, *A* is the presence of some amoeba in the stomach, and *B* is the existence of damage to the stomach lining. Identifying a correlation between of *A* and *B* tells you that the two are likely to be causally related but does not tell you in what way. Perhaps this amoeba causes stomach lining damage; perhaps stomach lining damage allows the amoeba to grow; or perhaps they share some other common cause. In the absence of available temporal or spatial cues,[1] the direction of causal connections can only be established by performing active *interventions* (experimental manipulations) of the variables. One can intervene, manipulating *A* and checking if this results in a change in *B* (or manipulating *B* and checking if this results in a change in *A*). In this example one might manipulate *A* by adding the amoeba to a sample of stomach wall tissue to see it becomes damaged or manipulate *B* by making cuts in a sample of stomach lining tissue to see if this results in growth of the amoeba. If manipulating *A* changes *B* then this is evidence that *A* is a cause of *B*. Thus, ability to intervene provides active learners with a privileged route for obtaining causal information.

Several studies have found that people benefit from the ability to perform interventions during causal learning (Coenen, Rehder, & Gureckis, 2014; Lagnado & Sloman, 2002, 2006, 2004; Schulz, 2001; Sobel, 2003; Sobel & Kushnir, 2006; Steyvers, Tenenbaum, Wagenmakers, & Blum, 2003). However, only Steyvers et al. and Coenen et al.'s studies explore how people select what intervention to perform, and both do so only for the case of a single intervention on a single variable in a semi-deterministic context. In contrast,

This article was published Online First October 20, 2014.

Neil R. Bramley, David A. Lagnado, and Maarten Speekenbrink, Department of Experimental Psychology, University College London.

Correspondence concerning this article should be addressed to Neil R. Bramley, Department of Experimental Psychology, University College London, 26 Bedford Way, London WC1H 0DS, United Kingdom. E-mail: neil.bramley.10@ucl.ac.uk

[1] It is true that for many everyday inferences there are strong spatial or temporal cues to causal structure. But, in other situations these cues are noisy, uninformative, or unavailable. Many systems propagate too fast to permit observation of time ordering of component activations (e.g., electrical systems), or they have hidden mechanisms (e.g., biological systems, psychological processes) or noisy/delayed presentation of variable values (Lagnado & Sloman, 2006). In others (e.g., crime scene investigation), observations come after the relevant causal process has finished. Additionally, people must be able to learn when spatial and temporal cues are reliable (Griffiths & Tenenbaum, 2009).

much real-world causal learning is probabilistic and incremental, taking place gradually over many instances. It has not yet been explored in what ways sequential active causal learning might be shaped by cognitive constraints on memory and processing or whether learners can plan ahead when choosing interventions.

Additionally, it has been shown that single-variable interventions are not sufficient to discriminate all possible causal structures (Eberhardt, Glymour, & Scheines, 2005; see also Figure 2 for an example). Interventions that simultaneously control for potential confounds to isolate a particular putative cause are cornerstones of scientific testing (Cartwright, 1989) and are key to scientific thinking (Kuhn & Dean, 2005). However, the only interventional learning study that allowed participants to perform multihold interventions was Sobel and Kushnir (2006), and this study did not analyze whether participants used these interventions effectively.

A final point is that no previous studies have explicitly incentivized causal learners. There is ambiguity in any assessment of intervention choices based on comparison to a correct or optimal behavior according to a single criterion value. This is because one cannot assume that the participants' goal was to maximize the quantity used to drive the analyses. In other areas of active learning research, researchers have run experiments to discriminate between potential objective functions that might underpin human active learning (e.g., Baron, Beattie, & Hershey, 1988; Gureckis & Markant, 2009; Meder & Nelson, 2012; Nelson, 2005; Nelson, McKenzie, Cottrell, & Sejnowski, 2010), but this is yet to be explored in the domain of causal learning.

Clearly, there are many aspects of active causal learning that call out for further exploration. Therefore, here we present two experiments and modeling that extend the existing work along several dimensions. In particular we explore

1. Whether people can choose and learn from interventions effectively in a fully probabilistic, abstract, and unconstrained environment.

2. To what extent people make effective use of complex "controlling" interventions as well as simple, single-variable fixes.

3. What objective function best explains participants' intervention choices: Do they act to maximize their expected utility, to maximize their probability of being correct, or to minimize their uncertainty?

4. Whether people choose interventions to learn in a stepwise, "greedy" way or whether there is evidence they can plan further ahead.

5. How people's causal beliefs evolve over a sequence of interventions. Is sequential causal learning biased by cognitive constraints such as forgetting or conservatism?

6. Whether people's interventions and causal judgments can be captured by simple heuristics.

The first two points can be addressed though standard analyses of participants' performances in various causal learning tasks. However, the latter questions lend themselves to more focused analyses of the dynamics of participants' intervention selections. Therefore, in the second half of the paper we will explore these questions by fitting a range of intervention and causal-judgment models directly to the actions and structure judgments made by participants in our experiments. We will compare different learning functions (utility gain, probability gain, and information gain); compare greedy learning models to models that plan ahead; and assess the influence of potential cognitive constraints (forgetting and conservatism). We will also explore the extent to which participants' behavior can be similarly captured by simple heuristic models as by more computationally complex Bayesian models.

Formal Framework

Causal Bayesian Networks

As have the authors of many recent treatments of causality in psychology, we take causal Bayesian networks (Pearl, 2000; Sprites, Glymour, & Scheines, 1993) as our starting point. We do not describe these in detail for space reasons, but see Heckerman (1998) for a tutorial on learning and using Bayesian networks for inference and Holyoak and Cheng (2011) for a review of recent applications of causal Bayesian networks in psychology.

Briefly, a Bayesian network (Pearl, 2000; Sprites et al., 1993) is a parameterized directed acyclic graph (see Figure 1). The nodes in the graph represent some variables of interest, and the directed links represent dependencies. Bayesian networks are defined by the Markov condition, which states that each node is independent of all of its nondescendants given its parents. Descendants are nodes that you can reach by following arrows from the current node, and parents are the nodes with arrows leading to the current node.

To capture causal structure, we interpret the directed links as causal connections going from cause to effect. In general, the causal variables can take any number of states or even be continuous, and the conditional probabilities of states of effect variables given cause variables can also take any arbitrary form. However, here we restrict ourselves to binary (present/absent) variables and causal connections that raise the probability of their effects. Following Cheng's (1997) power PC formalism and the noisy-OR combination function (Pearl, 1988), we assume the probability of an effect $e \in [1 = \text{Present}, 0 = \text{Absent}]$ given the presence of one or more causes $c_j \in [1 = \text{Present}, 0 = \text{Absent}]$ and some probability s of e spontaneously activating[2] is given by

$$p(e = 1 \mid c_1 = 1, \ldots, c_n = 1) = 1 - (1 - s)\prod_{j=1}^{n}(1 - power_{c_j \to e}) \tag{1}$$

where the *power* of each cause is defined as its probability of bringing about the effect in the absence of any other causes:

$$power_{c \to e} = \frac{p(e = 1 \mid c = 1) - p(e = 1 \mid c = 0)}{1 - p(e = 1 \mid c = 0)} \tag{2}$$

This captures the idea that the more (independently acting) causes are present, the more likely the effect.

[2] One can think of s as the combined *power* of the potential causes of e that are outside of the scope of the causal network. This means that even if there are no causes of effect e active in the network, there is still some chance of e occurring.

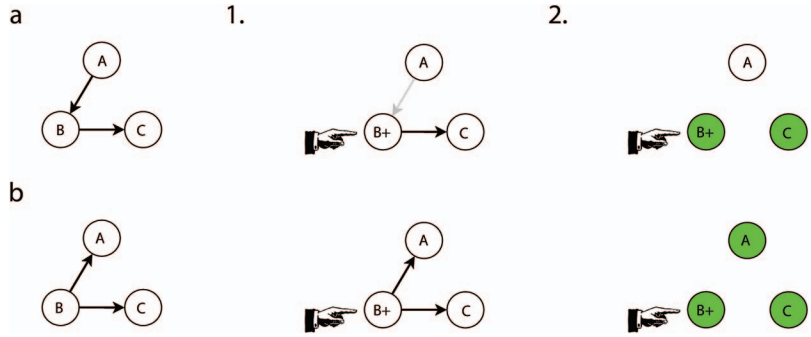

Figure 1. Two possible structures (a) $A \rightarrow B \rightarrow C$ and (b) $A \leftarrow B \rightarrow C$. 1. Intervening on B breaks any incoming causal links (dotted in gray). 2. Assuming the goal is to distinguish between these two structures, intervening on B is informative. Assuming binary variables and generative causal links, if we clamp B on (+ symbol) and then observe that C also turns on (highlighted in green [gray]) but A does not, then we have evidence for structure (a) $A \rightarrow B \rightarrow C$. If A also comes on, this is evidence for structure (b) $A \leftarrow B \rightarrow C$. See the online article for the color version of the figure.

To some extent, the causal structure likely to connect a set of variables can be identified through passive observation. Given data D and a prior over a set G of causal Bayesian network structures, one can compute the posterior probability of each network structure using known network parameters or by integrating over parameterizations θ. In the latter case, the posterior probability of a graph $g \in G$ is given by

$$p(g \mid D) = \frac{\int_\theta p(D \mid g, \theta) p(\theta \mid g) p(g) d\theta}{\sum_{g' \in G} \int_\theta p(D \mid g', \theta) p(\theta \mid g') p(g') d\theta} \quad (3)$$

However, a large amount of data compared to the number of possible causal networks is typically required to achieve this (Griffiths & Tenenbaum, 2009), and the number of possible networks rapidly becomes very large (25 for 3 variables, 543 for 4 variables, 29,281 for 5 variables, etc.), making learning extremely costly. Worse, without known parameters, many networks cannot be distinguished purely on the basis of covariational information (Pearl, 2000). Covariational information only allows a learner to identify that the network falls in a subset of possible causal networks, known as a Markov equivalence class, but not decide which network out of this subset is correct. This suggests that we must look beyond the calculus of observational learning to explain people's remarkable ability to learn the causal structure of their environment (Griffiths & Tenenbaum, 2009).

Interventions

The information provided by interventions is qualitatively different from that provided by observations. This is because intervened-on variables no longer tell you anything about their normal causes. This makes intuitive sense (i.e., turning your burglar alarm on tells you nothing about whether there has been a burglary, but attempting to break into your house may help you find out if your burglar alarm is working). In the causal Bayesian network framework, interventions can be modeled as *graph surgery* (Pearl, 2000; see Figure 1). Any links going to an intervened-on variable are temporarily severed or removed. Then, the variable is set, or *clamped,* to a particular value (i.e., clamped on or off in the binary case). The resulting values of the unclamped variables are then observed and inference is performed as normal. Intervening thus allows the learner to sidestep Markov equivalence issues and reduces the complexity of learning by reducing the number of variables that must be marginalized over with each new datum. To distinguish certain structures, one must clamp multiple variables at a time, as we do when controlling for potential confounds in an experiment. This is the case whenever you want to rule out direct connections between variables that you know are indirectly connected in a chain, as in Figure 2. With single-variable interventions these structures can only be distinguished on rare and lucky occasions (e.g., if the middle link in the chain fails but the direct link to the end variable still works). If the causal connections are perfectly reliable, a double hold intervention (where the middle link in the chain is clamped off) is necessary. Because the number of possible direct links increases at a rate of $\frac{(N-1)(N-2)}{2}$, where N is the length of the chain, the need for controlled interventions for causal learning about realistically complex problems is likely to be large. Eberhardt et al. (2005) show that, to identify any causal structure connecting N variables, at most $log_2(N) + 1$ idealized experiments are sufficient, provided any subset of variables can be clamped in each experiment. This proof assumes that each experiment is sufficiently highly powered (e.g., a random controlled trial or experimental manipulation with sufficient sample size) to reliably rule various structures out. However, in this paper we are interested in single-sample experiments (i.e., those typically achievable by real-world learners interacting with a causal system). Therefore, unless causal links are deterministic, a larger number of interventions are required to identify the correct causal structure with a high probability; at most, $log_2(N) + 1$ different types of intervention will be sufficient.

Quantifying Interventions

How can we quantify how useful an intervention is to a learner? Intuitively, good interventions provide evidence about aspects of

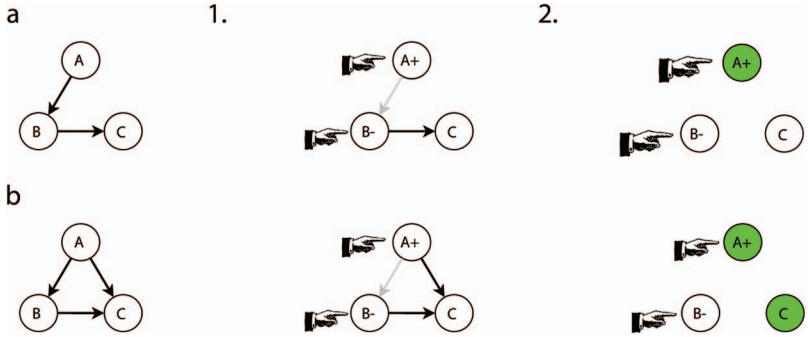

Figure 2. To distinguish (a) $A \rightarrow B \rightarrow C$ from (b) $A \rightarrow B \rightarrow C$, $A \rightarrow C$, you can manipulate A while simultaneously holding B constant. 1. This is achieved by clamping A on (+ symbol) and clamping B off (− symbol). 2. Then, if C still activates (highlighted in green [gray]), this is evidence for the second, fully connected structure. See the online article for the color version of the figure.

the structure of a causal system about which the learner is currently uncertain. This means that an intervention's value is relative to the learner's current uncertainty, as captured by a prior probability distribution over possible causal structures. Additionally, usefulness is necessarily relative to some goal, so we must also have a conception about what the learner is aiming to gain from an intervention. Finally, the effect of an intervention depends on its outcome, which is necessarily unknown at the time of its selection. Therefore, we must consider the expected usefulness of an intervention by summing or integrating over its possible outcomes and their likelihoods.

Mathematically, at a time point t, the marginal probability of each possible outcome $o \in O$ of an intervention $q \in Q$ and prior distribution over graphs $p_t(g)$[3] is given by

$$p_t(o \mid q) = \sum_{g \in G} p(o \mid q, g) p_t(g) \tag{4}$$

The posterior probability distribution over graphs, given an intervention–outcome pair, is then

$$p_t(g \mid q_t, o_t) = \frac{p(o_t \mid q_t, g) p_t(g)}{\sum_{g' \in G} p(o_t \mid q_t, g') p_t(g')} \tag{5}$$

These posteriors can be used to compute any desired summary values $V_t(G|o_t, q_t)$. For discrete outcomes, the *expected value* of an intervention is the sum of these values weighted by the probabilities of the different outcomes

$$E_o[V_t(G \mid o, q)] = \sum_{o \in O} V_t(G \mid o, q) p_t(o \mid q). \tag{6}$$

An optimal intervention $q_t^* \in Q$ can then be defined as the intervention for which the expected value is maximal

$$q_t^* = \arg \max_{q \in Q} E_o[V_t(G \mid o, q)]. \tag{7}$$

Choosing queries or experiments that will, in expectancy, maximize some sensible value of one's posterior is a cornerstone of Bayesian optimal experimental design and decision theory (Good, 1950; Lindley, 1956) but has been little explored in the context of interventional causal learning. Three commonly used measures in the active learning literature are (*expected) utility gain* (Gureckis & Markant, 2009; Meder & Nelson, 2012), *probability gain* (Baron, 2005), and *information gain* (Shannon, 2001; Steyvers et al., 2003; see Figure 3).

Utility gain. If you know how valuable correctly identifying all or part of the true causal system is, then the goal of your interventions is to get you to a state of knowledge about the true graph that is worth more to you than the one you were in before. Mathematically, this means maximization of your expectancy about your post-outcome, post-classification expected utility $E_o[U_t(G \mid q, o)]$.

If each potential graph judgment g has a utility given that the true graph is g', we can capture the value of any judgment by some reward function R (see Figure 3 for a simple example that also reflects the utilities of causal judgments in these experiments). Assuming one will always choose the causal structure with the highest expected reward, the *utility gain,* or $U(G)_{\text{gain}}$, of an intervention's outcome is the maximum over expected utilities of the possible judgments given the posterior $p_t(g' \mid q,o)$ minus the maximum for the prior $p_t(g')$:

$$U_t(G \mid q, o)_{gain} = \max_{g \in G} \sum_{g' \in G} R(g, g') p_t(g' \mid q, o) - \max_{g \in G} \sum_{g' \in G} R(g, g') p_t(g') \tag{8}$$

An optimal intervention is defined as the intervention that maximizes the expected utility gain (i.e., replacing V by U in Equations 6 and 7).

Probability gain. Although maximizing expected utility can be seen as the ultimate goal of intervening, often a useful proxy is to maximize your expected probability of being correct. Under many normal circumstances, choosing the most probable option will correspond to choosing the option that maximizes your expected utility (Baron et al., 1988); however, in terms of favoring one potential posterior distribution over another, the two values are more likely to differ depending on the reward function (see Figure

[3] Likelihoods $p(o|q, g)$ can be calculated using the known causal powers of the variables, or replaced with an integration over parameter values: $\int_\theta p(o|q, g, \theta) p(\theta|g) d\theta$ in the general case.

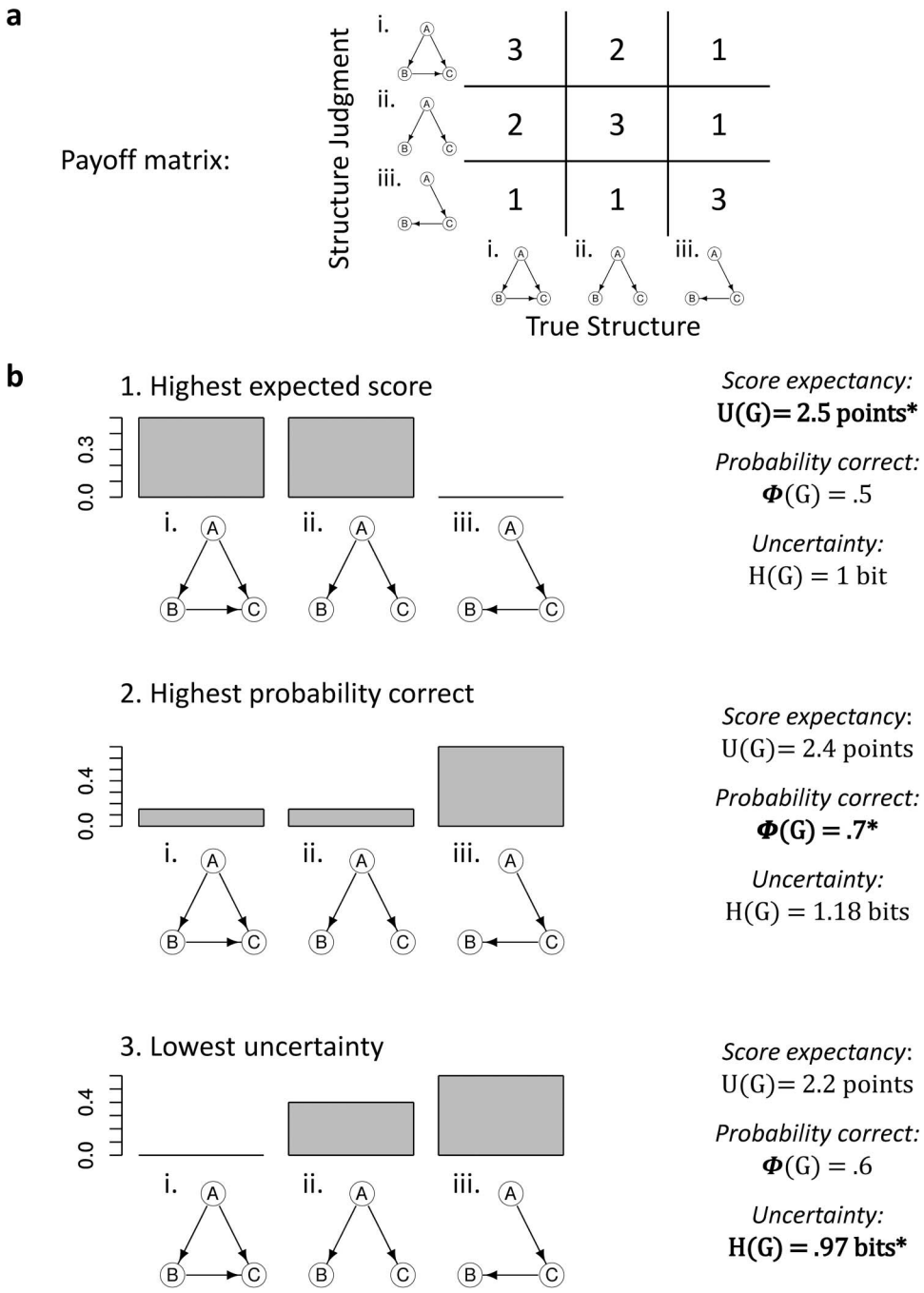

Figure 3. An example of differences in evaluation of posterior distributions with expected utility, probability correct, and uncertainty. (a) If the learner is paid one point per correctly identified connection then misclassifying structure i [$A \rightarrow B$, $A \rightarrow B$, $B \rightarrow C$] as structure ii [$A \rightarrow B$, $A \rightarrow C$] is less costly than misclassifying it as iii [$A \rightarrow C$, $C \rightarrow B$], because structures i and ii are more similar. (b) This makes Posterior 1 the most valuable in terms of maximizing expected payout. Posterior 2 has the highest probability of a completely correct classification, but uncertainty across the three structures is lowest overall in Posterior 3. U(G) = utility gain; ϕ(G) = probability gain; H(G) = information gain.

3). Assuming you will choose the causal structure that is most probable, the *probability gain,* or $\Phi(G)_{gain}$, can be written as

$$\Phi_t(G \mid q, o)_{gain} = \max_{g \in G} p_t(g \mid q, o) - \max_{g \in G} p_t(g) \quad (9)$$

An optimal intervention is defined as the intervention that maximizes the expected probability gain (i.e., replacing V by Φ in Equation 6 and 7).

Information gain. Another possible option for evaluating interventions comes from Shannon entropy. Shannon entropy (Shannon, 2001) is a measure of the overall uncertainty implied by a probability distribution. It is largest for a uniform distribution and drops toward zero as that distribution becomes more peaked. We can call reduction in Shannon entropy *information gain* (Lindley, 1956) and use this as a way to measure the extent to which a posterior implies a greater degree of certainty across all hypotheses, rather than just improvement in one's post-decision utility or probability of making a correct classification. Information gain, or $H(G)_{\text{gain}}$, is given by

$$H_t(G \mid q, o)_{gain} = \left[-\sum_{g \in G} p_t(g) \log_2 p_t(g) \right] - \left[-\sum_{g \in G} p_t(g \mid q, o) \log_2 p_t(g \mid q, o) \right] \quad (10)$$

An information-gain optimal intervention is defined as the intervention that maximizes the expected information gain (i.e., replacing V by H in Equation 6 and 7).

How do the measures differ? The extent to which these measures predict different intervention choices is one topic of investigation in this paper. However, as a starting point we can consider what types of posterior distribution they favor (see Figure 3). In the tasks we investigate here, people are rewarded according to how accurate their causal judgment is (e.g., how many of the causal connections and absences they correctly identify). This means that, according to expected utility, being nearly right is better than being completely wrong. Accordingly, we can expect that utility gain will favor interventions that divide the space of likely models into subsets of similar models rather than subsets of more diverse ones (see Figure 3b, No. 1). Probability gain is only concerned with interventions likely to raise the probability of the most likely hypothesis and does not care about similarity or overlap between hypotheses, or whether uncertainty between the various less probable options is reduced. Thus, we expect probability gain to favor interventions that are targeted toward confirming or disconfirming the current leading hypothesis. In contrast, information gain concerns the reduction in uncertainty over all hypotheses. It will favor interventions that are expected to make a large difference to the spread of probability across the less probable networks, even when this will not pay off immediately for the learner in terms of increasing utility or probability of a correct classification.

In support of the idea that probability gain might drive human active information search, Nelson, McKenzie, Cottrell, and Sejnowski (2010) has found participants' queries in a one-shot active classification task to be a closer match to probability gain than information gain. On the other hand, Baron et al.'s (1988) studies suggest that people will often select the question that has the higher information gain even if, for all possible answers, it will not change their resulting decision. There is also some recent evidence that people pick queries that are efficient in terms of information gain rather than utility gain in other areas of active learning (Gureckis & Markant, 2009; Meier & Blair, 2013). Steyvers et al. (2003) used information gain to quantify the intervention chosen by participants in their task, but they did not compare this with other measures. For these reasons, when analyzing our tasks we will consider utility gain, probability gain, and information gain alongside one another, asking to what extent the measures imply distinct patterns of interventions, and to what extent people's active causal learning choices appear to be driven by one or other measure.

Greedy or global optimization? A final issue is that, when learning continues over multiple instances, greedily choosing interventions that are expected to obtain the best results at the next time point (whether in terms of information, highest posterior probability, or expected utility) is not guaranteed to be optimal in the long run. There may be interventions that are not expected to give good results immediately but that provide the best results later on when paired with other interventions. To be truly optimal, a learner should treat each learning instance as a step in a Markov decision process (Puterman, 2009) and look many steps ahead, always selecting the intervention that is the first step in the sequence of interventions that leads to the greatest expected final or total utility (assuming the learner will maximize on all future interventions). However, computing expectancies over multiple hypotheses and interventions when each intervention has many possible outcomes is computationally intractable (Hyafil & Rivest, 1976) for all but the smallest number of variables and most constrained hypothesis spaces, dooming any search for strict optimality in the general case. It is an open question, which we will explore here, whether people can think more than one step ahead when planning interventions.

Experiment 1

Method

Participants. Seventy-nine participants from Mechanical Turk completed Experiment 1. They were paid between \$1 and \$4 (M = \$2.80), depending on performance.[4]

Design and procedure. To test people's ability to learn causal structure through intervention, we designed an interactive computer-based active learning task in Flash (see Figure 4; also see www.ucl.ac.uk/lagnado-lab/neil_bramley for a demo). In the task, participants had to use interventions to find and mark the causal connections in several probabilistic causal systems.

Participants completed one practice problem and five test problems. The practice problem was randomly chosen from the five test problems. The test problems were presented once each, in a randomized order. Participants performed 12 tests per problem as described below before finalizing their structure judgment and receiving a score.

For each problem, participants were faced with three filled gray circles, set against a white background. They were trained that

[4] The number of participants was purely dependent on our experimental budget of \$300. Unfortunately, age and gender were not properly stored for these participants.

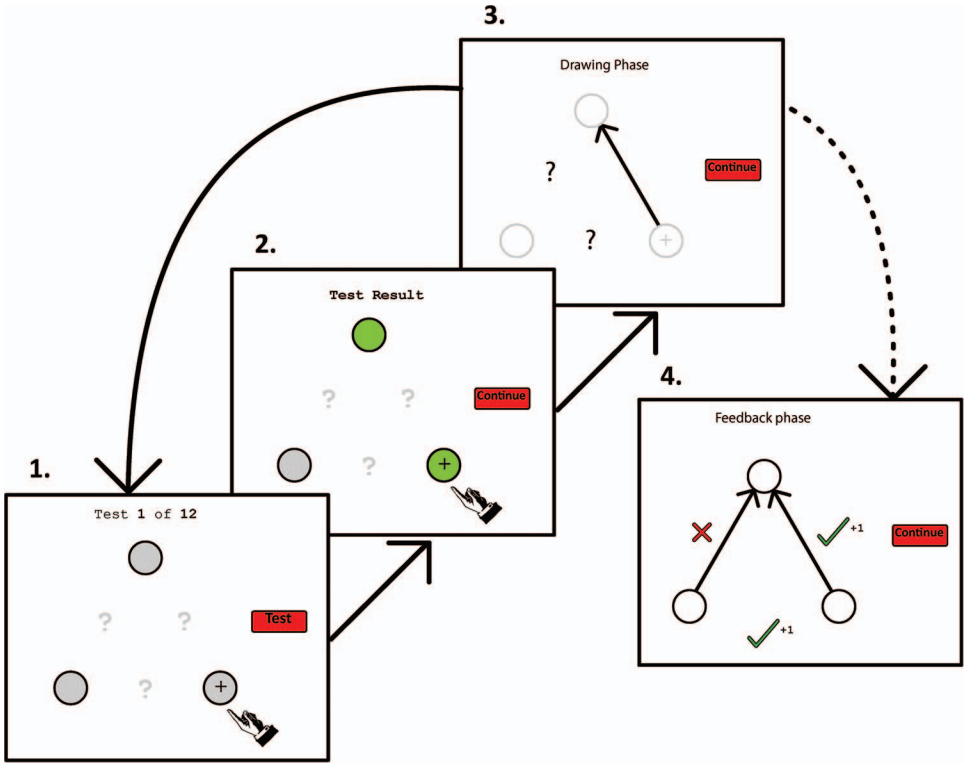

Figure 4. The procedure for a problem. 1. Choosing an intervention. 2. Observing the result. 3. Updating causal links. After 12 trials, 4. Getting feedback and a score for the chosen graph. See the online article for the color version of the figure.

these were nodes and that the nodes made up a causal system of binary variables, but they were not given any further cover story. Initially all of the nodes were inactive, but when participants performed a test then some or all of the nodes could temporarily activate. An active node glowed green and wobbled from side to side, while an inactive one remained gray. For each structure participants would perform multiple tests before endorsing a causal structure and moving on to the next problem. The running score, test number, and problem number were displayed across the top of the screen during testing. The location of nodes *A*, *B* and *C* were randomized, and the nodes were not labeled.

Each test had three main stages (see Figure 4).

1. First participants would select what intervention to perform. They could clamp between 0 and 3 of the nodes either to *active* or *inactive*. Clicking once on a node clamped it to active (denoted by a plus symbol), and clicking again clamped it to inactive (denoted by a minus symbol). Clicking a third time unclamped the node again. A pointing hand appeared next to clamped nodes to make it clear that they had been fixed by the participant.

2. Once the participant was happy with the intervention they had selected, he would press "Test" and observe the outcome of their test. The outcome would consist of 0–3 of the nodes activating. Whether a node activated on a given trial depended on the hidden causal connections and the choice of intervention. Participants were trained that nodes activated by themselves with a probability of .1 (unless they had been clamped, in which case they would always take the state they had been clamped in). They were also trained that causal links worked 80% of the time.[5] Therefore, clamping a node to active tended to cause any children of that node to activate, and this would tend to propagate to (unclamped) descendants. The noise in the system meant that sometimes there were false positives, where nodes activated without being caused by any of the other nodes, and false negatives, where causal links sometimes failed to work. The pattern of data seen by a participant over the task was thus a partly random function of the participant's intervention choices.

3. After each test there was a drawing phase in which participants registered their best guess thus far as to the causal connections between the nodes. Initially there was a question mark between each pair of nodes, indicating that no causal link had been marked there yet. Clicking on these question marks during the drawing phase would remove them and cycle through the options *no link, clockwise link, anticlockwise link,* back to *no link*. The initial direction of each link (clockwise or anticlockwise) was randomized. Participants were not forced to mark or update links until after the final test but were invited to mark as they went along as a memory aid. This approach was used to avoid forcing participants to make specific judgments before they had seen enough

[5] Concretely, they had a causal power of.8. Combining causal power with the spontaneous activation rate, a node with one active cause had a $1 - (1 - .1)(1 - .8) = .82$ probability of activating.

information to make an informed judgment and to maximize our record of their evolving judgment during the task.

4. Participants performed 12 tests on each problem. After their last test, they were prompted to finalize their choice for the causal structure (i.e., they had to choose *no link, clockwise link,* or *anticlockwise link* for all three pairs of nodes, leaving no question marks). Once they had done this they were given feedback as to the correct causal structure and received one point for each correctly identified link (see Figure 4). There were three node pairs per problem (A-B, A-C, and B-C) and three options (*no-link, clockwise link, anticlockwise link*) per node pair. This means that chance-level performance was 1 correct link per problem, or ≈5 points over the five problems, while an ideal learner could approach 15 points. At the end of the task participants received $1 plus 20 cents per correctly identified link, leading to a maximum payment of $4.

Before starting the practice round, participants completed a comprehensive and interactive instructions section designed to familiarize them with the spontaneous activation rate, the causal power of the nodes, the role of the different interventions, and the aim of the task (see demo at www.ucl.ac.uk/lagnado-lab/neil_bramley). To train participants on the causal power of these connections, we presented them with a page with five pairs of nodes. The left node of each pair was clamped on, and it was revealed that there was a causal connection from each left node to each right hand (unclamped) node. Participants were made to test these networks at least 4 times, finding that an average of 4/5 of the unclamped nodes would activate. The outcomes of their first three tests were fixed to reflect this probability, and thereafter the outcomes were generated probabilistically. Similarly, for the rate of spontaneous activations, participants were made to perform at least four tests on a page full of 10 unclamped and unconnected nodes, where an average of 1/10 of these would activate on each test. In addition to this experience-based training, participants were told the probabilities explicitly. Before starting the task they had to answer four multiple choice questions checking they had understood: The goal of the task (e.g., how to win money); The role of clamping variables on and the role of clamping them off; and The probabilistic nature of the networks. If the participant got less than 3 of 4 questions correct they were sent back to the beginning of the instructions.

Results

Participants identified an average of 9.0 out of 15 (SD = 4.1) causal links and got 34% of the models completely right. This is well above the chance level of 5 out of 15 correct links (and 3.7% models correct), $t(78) = 8.60$, $p < .0001$. However, the distribution of performance appears somewhat bimodal with one mode at chance and the other near ceiling (see Figure 5), suggesting that some participants were not able to solve the task while others did very well. This bimodality is confirmed by a dip test (Hartigan & Hartigan, 1985), $D = 0.09$, $p < .0001$. There was no effect of problem order on performance, $F(1, 394) = 0.06$, $\eta^2 = 0$, $p = .81$, nor did participants perform better on the problem they faced as their practice trial and when they faced it again as a test problem, $t(110) = -1.12$, $p = .26$. Participants did not overconnect or underconnect their final causal structures, on average opting for no-link for 30% of node pairs, which was very close to the true percentage of 33%.

Participants were about equally accurate on the different structures, with slightly lower scores for the chain, common cause, and

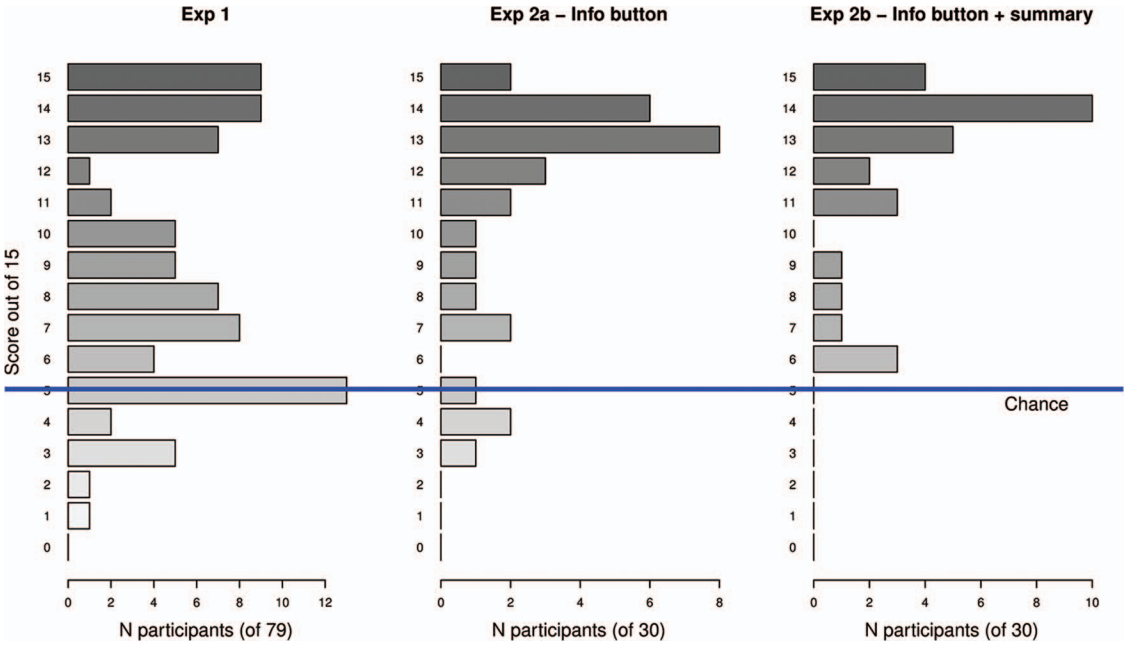

Figure 5. Histogram of scores in Experiment (Exp) 1 and Experiments 2a and 2b. There were 15 points available in total (identifying all 15 connection-spaces correctly), and one could expect to get an average of 5 of these right by guessing. See the online article for the color version of the figure.

fully connected structures than for the common effect or singly connected, but there was no main effect of problem on score, $F(4, 390) = 0.87$, $\eta^2 = .01$, $p = .48$. However, looking at modal structure judgment errors, one error stands out dramatically: 18 participants mistook the chain [$A \rightarrow B$, $B \rightarrow C$] for the [$A \rightarrow B$, $A \rightarrow C$, $B \rightarrow C$] fully connected structure. This was almost as many as those participants who correctly identified the structure (see Table 1).

Experiment 2

Before Experiment 1 is analyzed further, an immediate question is why there was so much variance in participants' performance. One explanation for this could be that there are important individual differences between participants that strongly affected their ability to learn successfully. Steyvers et al.'s (2003) modeling suggested that people's ability to remember evidence from multiple past trials may be a critical psychological bottleneck for active causal learning. One way to check if poor performance stems from an inability to remember past tests is to provide participants with a history of their past interventions and their outcomes and assess whether this leads to better and more consistent causal learning. However, another, perhaps simpler explanation for the variance is that some participants were confused about what to do and so responded randomly for all or much of the experiment.

To test both of these explanations, we ran another experiment using the same task as in Experiment 1 but with two additions. In Experiment 2a we provided an information button, which would bring up a text box reminding them about what they were supposed to do at that stage of the task. In Experiment 2b, participants were still provided with this information button, but in addition they were provided with a summary of all their past tests and their outcomes for the current problem. These were shown in a 4×3 grid to the left of the screen. After each test a new cell would be filled with a picture showing the causal system, the interventions selected (marked with plus and minus symbols as in the main task), and the nodes that activated (shown in green [dark gray] as in the main task; see Figure 6).

Method

Participants. Sixty additional Mechanical Turk participants aged 18 to 64 ($M = 31.4$ years, $SD = 11.2$) completed experiment 2. Once again, participants were paid between \$1 and \$4 ($M =$ \$3.32, $SD = .65$).

Design and procedure. The procedure was exactly as in Experiment 1, except that now half of the participants were randomly assigned to Experiment 2a (info button only) and the other to Experiment 2b (info button + summary).

Results

On average, judgment accuracy in Experiment 2 was considerably higher than in Experiment 1, $t(136.7) = 4.2608$, $p < .0001$. Participants in Condition 2a (info button only) scored significantly higher at 11.1 (out of 15) correct links ($SD = 3.5$) than those in Experiment 1, $t(60.1) = 2.7$, $p = .009$, while participants in Condition 2b (info button + summary) were slightly higher at 12.13 ($SD = 2.9$), again significantly higher than in Experiment 1, $t(74.2) = 4.5$, $p < .0001$. However, the improvement from 2a to 2b was not significant, $t(55.7) = 1.238$, $p = .22$. Inspecting Figure 5, we see that the number of participants performing close to chance is greatly reduced in both Experiment 2 conditions compared to Experiment 1, accounting for this difference in average performance.

These differences suggest that many of the poorer performers in Experiment 1 were simply confused about the task rather than being particularly poor at remembering evidence from past trials. However, scores were so high in Experiment 2 that failure to detect a performance-level difference between conditions may be partly due to a ceiling effect. In line with this, we see that participants in Experiment 2b (info button + summary) were significantly faster at completing the task, at 18.4 min ($SD = 8.1$), than were those in the Experiment 2a (info button only condition), at 24.3 min ($SD =$ 12.1), $t(50.6) = 2.26$, $p = .03$. This suggests that the summary made a difference in terms of the effort or difficulty of at least some aspects of the task.

As in Experiment 1, there was no main effect of causal network type on performance in Experiment 2, $F(4, 295) = 0.64$, $\eta^2 = .008$, $p = .63$, nor were there any significant interactions between performance on the different structures and whether participants saw summary information (all ps $> .05$). However, as in Experiment 1, we found that participants were very likely to add a direct $A \rightarrow C$ connection for the chain structure (see Figure 7 and Table 1).

Table 1
The Three Most Frequent Causal Judgment Errors for Experiments 1, 2a, and 2b

Experiment	True structure	*N* correct	Mistaken for	*N* error
1	chain$^{[A \rightarrow B \rightarrow C]}$	20 (25%)	fully connected$^{[A \rightarrow B \rightarrow C,\ A \rightarrow C]}$	18 (23%)
1	chain$^{[A \rightarrow B \rightarrow C]}$	20 (25%)	common cause$^{[B \leftarrow A \rightarrow C]}$	7 (9%)
1	fully connected$^{[A \rightarrow B \rightarrow C,\ A \rightarrow C]}$	26 (35%)	chain$^{[A \rightarrow B \rightarrow C]}$	7 (9%)
2a	chain$^{[A \rightarrow B \rightarrow C]}$	12 (40%)	fully connected$^{[A \rightarrow B \rightarrow C,\ A \rightarrow C]}$	12 (40%)
2a	common effect$^{[A \rightarrow B \leftarrow C]}$	17 (56%)	fully connected$^{[C \rightarrow A \rightarrow B,\ C \rightarrow B]}$	3 (9%)
2a	common cause$^{[B \leftarrow A \rightarrow C]}$	15 (50%)	fully connected$^{[C \rightarrow B \rightarrow A,\ C \rightarrow A]}$	3 (9%)
2b	chain$^{[A \rightarrow B \rightarrow C]}$	18 (60%)	fully connected$^{[A \rightarrow B \rightarrow C,\ A \rightarrow C]}$	8 (26%)
2b	fully connected$^{[A \rightarrow B \rightarrow C,\ A \rightarrow C]}$	17 (56%)	chain$^{[A \rightarrow B \rightarrow C]}$	4 (13%)
2b	common effect$^{[A \rightarrow B \leftarrow C]}$	21 (70%)	fully connected$^{[C \rightarrow A \rightarrow B,\ C \rightarrow B]}$	3 (9%)

Note. *N* correct is the number of participants who identified this structure correctly, and *N* error is the number of participants to make this particular error.

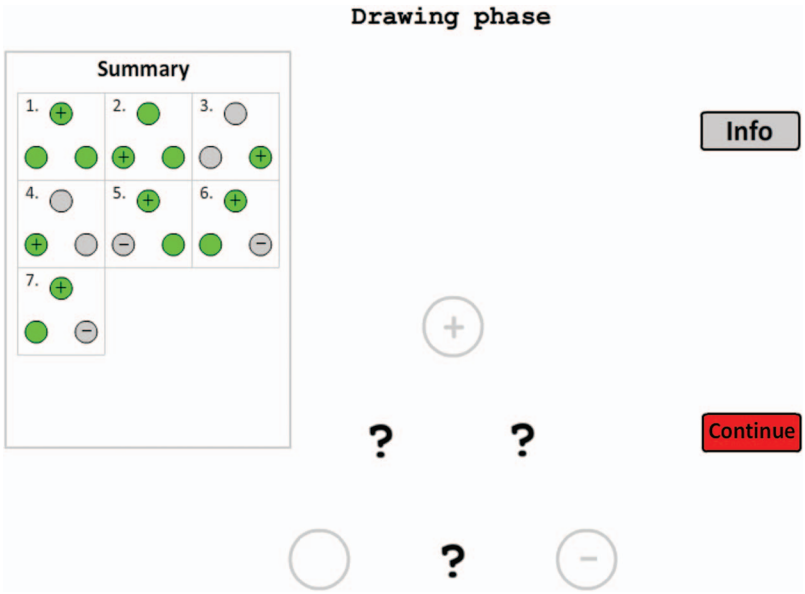

Figure 6. Interface in Experiment 2. Note the info button on the right (2a and 2b) and the summary information provided on the left (2b only: Nodes with a + symbol were clamped on, those with a − symbol were clamped off, and those with no symbol were unclamped. Green [dark gray] nodes activated and gray [light gray] ones did not). See the online article for the color version of the figure.

We now move on to analyze the interventions selected by participants in the two experiments. Because the task in Experiments 1 and 2 was fundamentally the same, we will mainly report analyses for all 139 participants together but where relevant will also explore differences between Experiment 1 and 2 and between the two conditions of Experiment 2.

Intervention Choices

Benchmarking the interventions. Participants' ultimate goal was to maximize their payout at the end of each problem (after their 12th test). However, as mentioned in the introduction, there are various approaches to choosing interventions expected to help achieve this goal. Here we use three "greedy" (one-step-ahead) value functions—expected utility, probability, and information gain—to assess how effectively participants selected different interventions.

To get a picture of the sequences of interventions favored by efficient utility, probability, or information seeking learners, we simulated the task 100 times using one-step-ahead expected utility, probability, and information gain (as defined in the introduction) to select each intervention. The prior at each time point was based on Bayesian updating from a flat prior using the outcomes of all previous interventions. All three measures always favored simple interventions $A+$, $B+$, or $C+$ (see Figure 8) for the first few tests for which the prior was relatively flat. Then, as they become more certain about the underlying structure, they increasingly selected controlled interventions with one node clamped on and another clamped off (e.g., $A+$, $B-$). After six tests, the probability that the models would select one of these controlled interventions was .41 for the utility gain model, .37 for probability gain model, and .51 for information gain model. For the later tests, if expected utility of the prior was already very close to 3 (full marks) and probability of correct classification was very close to 1, probability and utility gain were unable to distinguish between interventions, assigning them all expected gains of zero. Whenever this happened, these models would select interventions randomly. Information gain meanwhile continued to favor a mixture of simple and controlled interventions. The information gain model would occasionally select an intervention with two nodes clamped on (4.5% of the time on tests 9 to 12). Other interventions (e.g., clamping two nodes off or clamping everything) did not provide any information about the causal structure so had expected gains of zero. These were selected only by the utility and probability gain models and only on the last few trials, when they could not distinguish between the interventions and so selected at random. The three approaches averaged scores of 14.1 (utility), 14.2 (probability), and 14.6 (information) correct links (see Figure 9). Thus, within 12 trials it was possible for an efficient one-step-ahead intervener to approach a perfect performance, averaging at least 14/15 depending on the choice of value function driving intervention choices.

Looking two steps ahead, the efficient active learners using one of these measures average almost identical average final scores (14.6, 14.4, and 14.6 points, respectively) despite still using somewhat different sequences on interventions. The two-step-ahead models selected a higher proportion of controlled interventions than the greedy models (38%/30% for information gain). Two-step-ahead probability and utility gain were always able to distinguish between the interventions, meaning they would no longer select interventions randomly on later tests.

For comparison, merely observing the system without clamping any variables would have provided very little information, capping

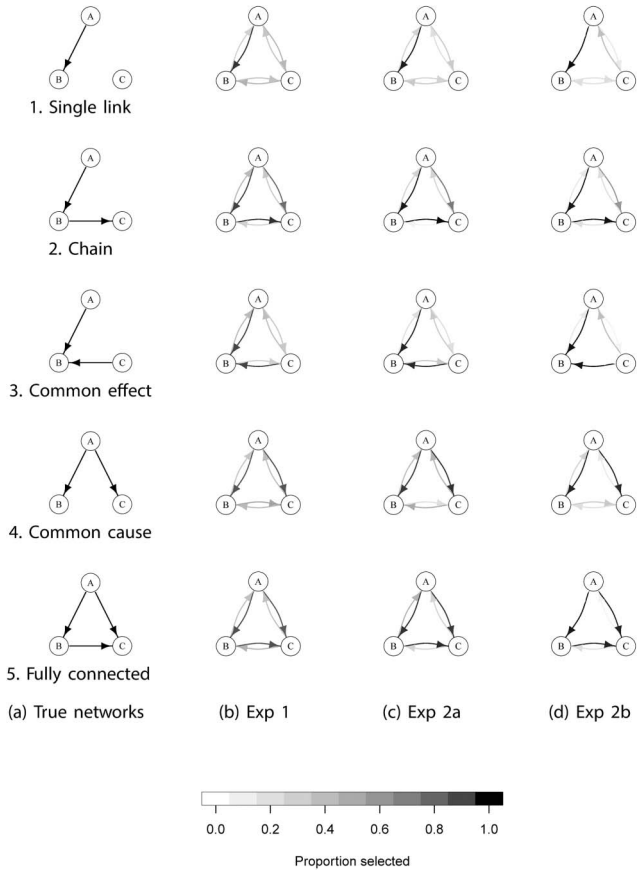

Figure 7. a. Causal test structures used in the experiments. From top to bottom: 1. A-B single link [$A \rightarrow B$]. 2. ABC chain [$A \rightarrow B$, $B \rightarrow C$]. 3. Common effect B [$A \rightarrow B$, $C \rightarrow B$]. 4. Common cause A [$A \rightarrow B$, $A \rightarrow C$]. 5. Fully connected ABC [$A \rightarrow B$, $A \rightarrow C$, $B \rightarrow C$]. b–d. Averaged final judgments over Experiments 1, 2a info button only and 2b info button + summary. Note that the direct link was often marked for the chain structure (row 2).

a learner's ideal score at an average of 1.87 points per problem (9.35 overall, or a .26 probability of identifying the correct graph).

Participant intervention choices.

Efficiency of intervention sequences. On average, participants selected highly efficient interventions in terms of utility, probability, and information gain. Participants learned much more than they would by picking interventions at random, and they selected interventions that put their achievable score much closer on average to the benchmark models than to a random intervener in terms of their final expected utilities (see Figure 9). Participants finished problems having learned enough that they could optimally score an average of 13.4 ($SD = 3.0$) points per problem ($M = 13.1$, $SD = 1.9$ in Experiment 1; $M = 13.5$, $SD = .35$ in Experiment 2) and have a .72 ($SD = .25$) probability of getting each graph completely right ($M = .70$, $SD = .25$ in Experiment 1, $M = .75$ $SD = .23$ in Experiment 2). This was significantly higher than selecting interventions at random, which would permit an average of only 11.6 points, or a .47 probability of getting each graph completely correct, $t(694) = 24$, $p < .0001$. However, it was still significantly lower than what could be achieved by consistently intervening to maximize utility, $t(694) = -12.7$, $p < .0001$, probability, $t(694) = -12.7$, $p < .0001$, or information gain, $t(694) = -18.3$, $p < .0001$. The quality of participants' interventions was strongly positively associated with their ultimate performance. This is true for all measures of intervention quality tested here: utility gain, $F(1, 137) = 63$, $\eta^2 = 0.31$, $p < .0001$; probability gain, $F(1, 137) = 81$, $\eta^2 = 0.37$, $p < .0001$; information gain, $F(1, 137) = 87$, $\eta^2 = 0.39$, $p < .0001$.

Simple interventions. As with the efficient learning models, simple interventions $A+$, $B+$, and $C+$ were by far the most frequently selected, accounting for 74% of all interventions despite constituting only 3 of the 27 selectable interventions (see Table 2). Propensity to use simple interventions was positively associated with performance across participants, $F(1, 137) = 41$, $\eta^2 = .23$, $p < .0001$. As with the efficient learner models, the probability a participant would select a simple intervention was highest at the start and then decreased over tests, $\beta = -.03 \pm .007$, $Z = -4.1$, $p < .0001$ (see Figure 8).

Controlled interventions. Controlled interventions (e.g., $A+$, $B-$) were selected only 7.4% of the time overall. This is not nearly as often they were selected by the efficient learner models. However, in line with these models, participants' probability of selecting a controlled intervention increased over tests, $\beta = .06 \pm .01$, $Z = 5.2$, $p < .0001$ (see Figure 8). Propensity to use controlled interventions was also positively associated with performance, $F(1, 137) = 14.1$, $\eta^2 = .09$, $p = .0002$. For each additional informative controlled intervention performed, participants scored .18 additional points in a task. The chain and fully connected structures are the two that cannot easily be distinguished without a controlled intervention (see Figure 2), and accordingly we find use of controlled interventions is higher when the generating causal network is a chain or fully connected structure, $\beta = 0.50 \pm .08$, $Z = 6.0$, $p < .0001$. In line with this, the use of controlled interventions also significantly predicts participants' probability of correctly omitting the $A \rightarrow C$ connection in the chain structure, $\beta = 1.7 \pm .4$, $Z = 4.6$, $p < .0001$. In addition, a higher proportion of participants used controlled interventions at least once in Experiment 2 than in Experiment 1, $\chi^2(60) = 9.3$, $p = .002$ (41/60 compared to 44/79).

The fact that participants performed fewer controlled interventions in later tests than the benchmark efficient learner models is consistent with the idea that they were slower to learn. This would mean they would require more of the simple interventions to reach a level of certainty under which controlled interventions become the most valuable choice. The modeling in the next section will allow us to explore this possibility.

Other interventions. Participants sometimes selected interventions with two nodes clamped on (e.g., $A+$ $B+$), doing so 10% of the time. While the information gain model would select these interventions occasionally in later trials, participants were just as likely to select them early on, $\beta = 0.009 \pm 0.01$, $Z = 0.775$, $p = .4$, and their propensity to select them was negatively associated with their performance, $F(1, 137) = 50$, $\eta^2 = .27$, $p < .0001$. This suggests that participants typically did not use these interventions efficiently, or did not learn from them appropriately. Frequency of clamping everything (e.g., $A+$, $B-$, $C-$) was strongly negatively correlated with performance, $F(1, 137) = 50$, $\eta^2 = .27$, $p < .0001$. Participants who selected this type of intervention averaged final

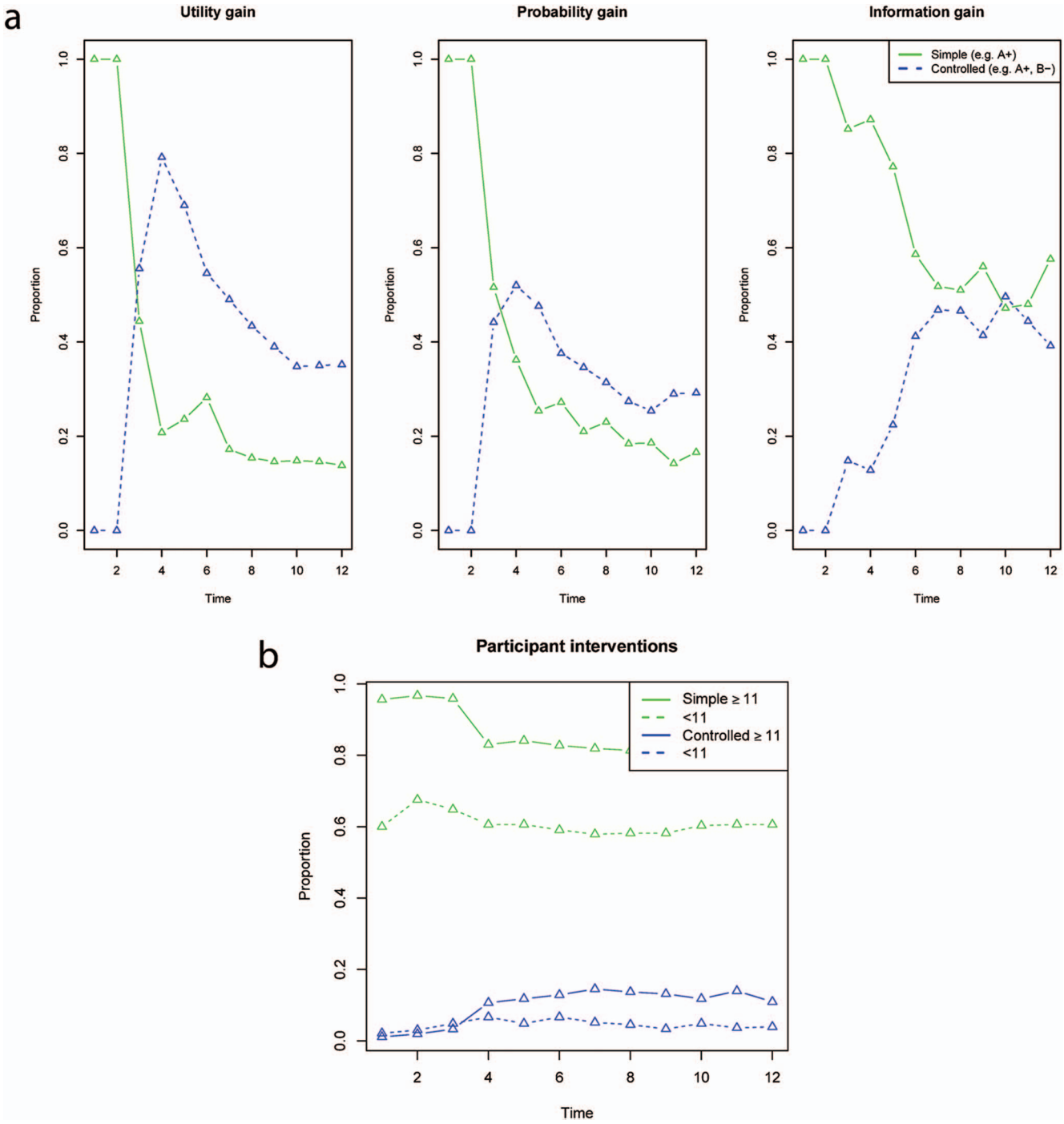

Figure 8. a. Proportion of simple (e.g., A+) versus controlled (A+, B−) intervention choices for the three efficient learning models averaged over 100 simulations of the task. For later tests, based on increasingly peaked priors, expected utility gain and probability gain no longer distinguish between interventions and start to choose randomly while information gain continues to distinguish. b. Participants' proportion simple and controlled interventions over both experiments with a median split by performance. See the online article for the color version of the figure.

scores of only 8. Observing with no nodes clamped and clamping one or two nodes off was rarely selected and was not significantly associated with performance (*p*s = .14, .06, and .91, respectively).[6]

Modeling Intervention Selection and Causal Judgments

So far, we have analyzed people's intervention selections at a relatively high level, looking only at how often particular types of intervention are chosen on average, either by good or bad participants, early or late during learning, or depending on the underlying causal structure. These high-level analyses have addressed the first two of our research questions, answering both in the positive:

1. The majority of people are able to choose informative interventions and learn causal structure effectively, even when the environment is fully probabilistic and abstract and there is a large space of causal structures.

2. Most people can make use of complex controlling interventions to disambiguate between otherwise hard-to-distinguish structures. Ability to do this is a strong predictor of correctly identifying the causal network, especially when the true network is a chain.

So far we have not touched upon the clear differences between interventions that fall within the same category (e.g., selecting $A+$ will provide very different information to $B+$ or $C+$ depending on the learner's current beliefs). Additionally, we have not yet tried to distinguish which intervention selection measure is more closely in line with participants' choices. Looking across all three experiments, the three value functions favor different intervention(s) to one another on many of participants' tests. Utility and probability gain disagree about what intervention should have been chosen on 19% of participants' tests. Utility and information gain disagree on 36% of participants' tests, and probability gain and information gain disagree 39% of participants' tests. However, simply counting the frequency of agreement between participants' interventions and those considered most valuable by one or other measure is a blunt instrument for understanding participants' actions. The measures do not just give a single favored intervention but give a distinct value for each of the 27 possible interventions. Furthermore, the benchmark models assume perfect Bayesian updating after each intervention while a richer model comparison should allow us to compare the different measures while relaxing the assumption that participants are perfect Bayesians. Thus, to progress further we will now fit and compare a range of models to participants' sequences of interventions and structure judgments. This will allow us to address our other research questions:

3. What objective function best explains people's choices.
4. Whether people can plan more than one intervention ahead.
5. Whether their belief update process is biased or constrained.
6. Whether we can capture their active learning with simple heuristics.

On each test a participant chooses an intervention but also can update their causal judgment by marking the presence or absence of possible causal links. The models discussed below will describe the intervention selections and causal judgments simultaneously, by assigning a probability to each intervention choice (from the 27

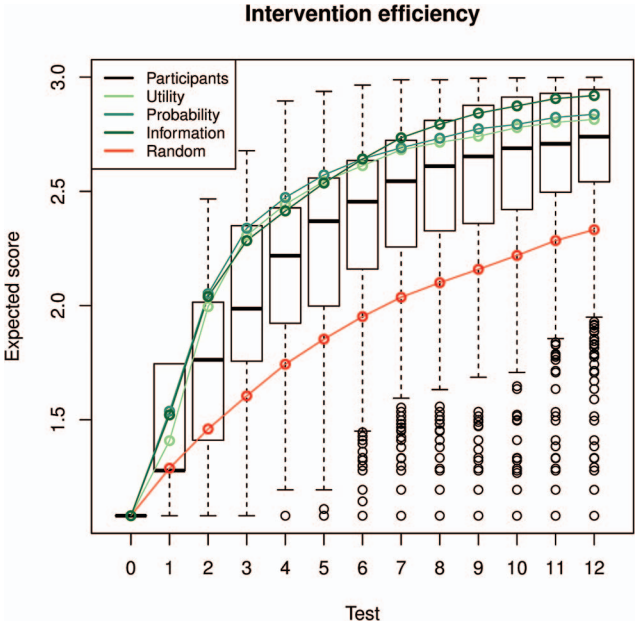

Figure 9. Participants' expected scores given the interventions participants had selected thus far. This is the best a participant could expect to score given the interventions and outcomes he or she had experienced up until that time point, averaged over the five problems. For comparison, the other lines denote the mean expected scores of expected utility, probability, or information maximizing active learners (shades of green [light gray]) and a passive learner who selects interventions at random (red [dark gray]), based on the simulations detailed in the text. See the online article for the color version of the figure.

Table 2
Comparison of the Proportion of Interventions of Different Types Selected by Participants in Experiments 1 and 2 (Conditions A and B), by Simulations Selecting Interventions at (R)andom, and by Maximizing Expected (U)tility, (P)robability, and (I)nformation Gain

	Proportion selected						
	Experiment			Simulation			
Intervention type	1	2a	2b	R	U	P	I
Observation	.01	.01	.02	.04	.08	.16	.00
Simple (e.g., $A+$)	.73	.73	.77	.11	.34	.38	.68
Controlled (e.g., $A+$, $B-$)	.06	.09	.10	.22	.41	.30	.30
Strange 1 (e.g., $A+$, $B+$)	.11	.08	.05	.11	.07	.08	.02
Strange 2 (e.g., $A-$)	.02	.02	.01	.11	.05	.05	.00
Strange 3 (e.g., $A-$, $B-$)	.00	.00	.00	.11	.05	.05	.00
Overcontrolled (e.g., $A+$, $B-$, $C-$)	.08	.07	.05	.30	.00	.00	.00

[6] Clamping off two nodes provides no information about the causal connections. Arguably, it still provides information about the spontaneous activation rate of variables, but participants had already been trained on this in the instructions.

legal interventions) and to each combination of marked and unspecified links (out of the 27 possible combinations of causal connections). Free parameters are fitted to individual data, because it is reasonable to assume that properties such as memory and learning strategy are fairly stable within subjects but likely to differ between subjects in ways that may help us understand what drives the large differences in individual performance.

We fit a total of 21 models (see Figure 10) separately to each participant's data. The models can be classified as either expectancy-based or heuristic models. The expectancy-based models assume that people choose interventions according to the expected value of each intervention, maximizing either utility, probability, or information gain (see Quantifying Interventions). The models assume that the expectancies, as well as causal judgments, are based on Bayesian updating of probability distributions over the causal structures and the models are rational in the sense that they are optimal with respect to people's goals, although we also allow for the possibility of cognitive constraints such as forgetting and conservatism.

In Steyvers et al.'s (2003) study, many participants chose models that suggested they remembered only the result of their final intervention (having apparently forgotten or discounted the evidence from their previous observations), while others seemed to remember a little more. This is in line with what we know about the limited capacity of working memory (Cowan, 2001; Miller, 1956) and its close relationship with learning (Baddeley, 1992). Thus it seems likely that people are somewhat "forgetful", or exhibit recency with respect to integrating the evidence they have seen. We expect that in Experiment 2b, where a summary of past outcomes is provided, memory load should be reduced and participants should display less recency effects.

With regard to conservatism, research suggests that people interpret new data within their existing causal structure beliefs wherever possible (e.g., Krynski & Tenenbaum, 2007). Anecdotally, people are typically slow or reluctant to change their causal beliefs. This suggests that people may also be *conservative* (Edwards, 1968) when updating their causal beliefs, even during learning. An additional motivation for this idea is the consideration that appropriate conservatism could actually complement forgetting; people may mitigate their forgetfulness about old evidence by remembering just what causal structural conclusions they have previously drawn from it (Harman, 1986). For example, suppose a participant registers an A → B causal link after the first three interventions. We can take this as a (noisy) indication he is fairly confident at this stage that, whatever the full causal structure is, it is likely to be one with a link from A to B. By the time the participant comes to his sixth intervention, he might not remember why he had concluded three trials earlier that there is an A → B link, but he would still be sensible to assume that he had a good reason for doing so at the time. This means that it may be wise to be conservative, preferring to consider models consistent with links you have already marked rather than those that are inconsistent even when you cannot remember why you marked the links in the first place.

In the heuristic models, intervention selections are not based on Bayesian belief updating and the expected value of interventions but are derived from simple rules of thumb. Although these models are not optimal with respect to any criterion, they can approximate the behavior of the rational models reasonably well.

Expectancy-Based Models

We will call a model that assumes participants are pure pragmatists, choosing each intervention with the goal of increasing their expected score, a *utilitarian* model. A utilitarian model assumes that participants choose interventions that are expected to maximize their payment at the next time point, or utility gain $U(G)$ (see Quantifying interventions).[7] We will call a model that assumes people are just concerned with maximizing their probability of being completely right (disregarding all other possible outcomes, or their payouts) a *gambler* model. A gambler model assumes participants choose actions that are expected to maximize the posterior probability of the most likely structure, or *probability gain* $\Psi(G)$. We will call a model that assumes people try to minimize their uncertainty (without worrying about their probability of being right, or how much they will get paid) a *scholar* model. A scholar model assumes that participants choose actions to maximize their expected *information gain,* $H(G)$, about the true structure at the next time point.

Updating causal beliefs and forgetting. All expectancy-based models assume that the learner's causal beliefs are represented by a probability distribution over all possible causal structures. At each time point, this probability distribution is based on Bayesian updating of their prior from the previous time point to incorporate the evidence provided by the outcome of their latest intervention. However, rather than a complete Bayesian updating (Equation 5), we allow for the possibility that evidence from past trials may be partly discounted or forgotten.

There are various ways to model forgetting (Lewandowsky & Farrell, 2010; Wixted, 2004). A reasonable (high-level) approach is to assume that people will forget random aspects of the evidence they have received, leading to a net "flattening" of participants' subjective priors going into each new intervention. We can formalize this by altering the Bayesian update equation, such that a uniform distribution is mixed with the participants' prior on each update to an extent controlled by a forgetting parameter $\gamma \in [0, 1]$. So instead of

$$p_t(g \mid q_t, o_t) \propto p(o_t \mid q_t, g) p_t(g)$$

as in Equation 5, we have

$$p_t(g \mid q_t, o_t) \propto p(o_t \mid q_t, g) \left[(1 - \gamma) p_t(g) + \gamma \frac{1}{m} \right] \tag{11}$$

where m is the total number of structures in G, and distributions are computed recursively as $p_t(g) = p_{t-1}(g|q_{t-1}, o_{t-1})$. By setting γ to 0 we get a model with no forgetting, and by setting it to 1 we get a model in which everything is forgotten after every test.

Choosing interventions. The expectancy-based models assume that intervention choices are based on the expected values of interventions. Let $v_{1t}, \ldots, v_{nt}$ denote the expected values $v_{qt} = E_o[V_t(G|q, o)]$, where the generic function V is identical to the

[7] For each judgment the expected payout was calculated as the points received for that judgment summed over every possible graph, each multiplied by the posterior probability of the graph. As an example, endorsing common cause [A → B; A → C] given the true structure is the chain [A → B; B → C] was worth one point because one of the three link-spaces (A-B) is correct and the other two are wrong.

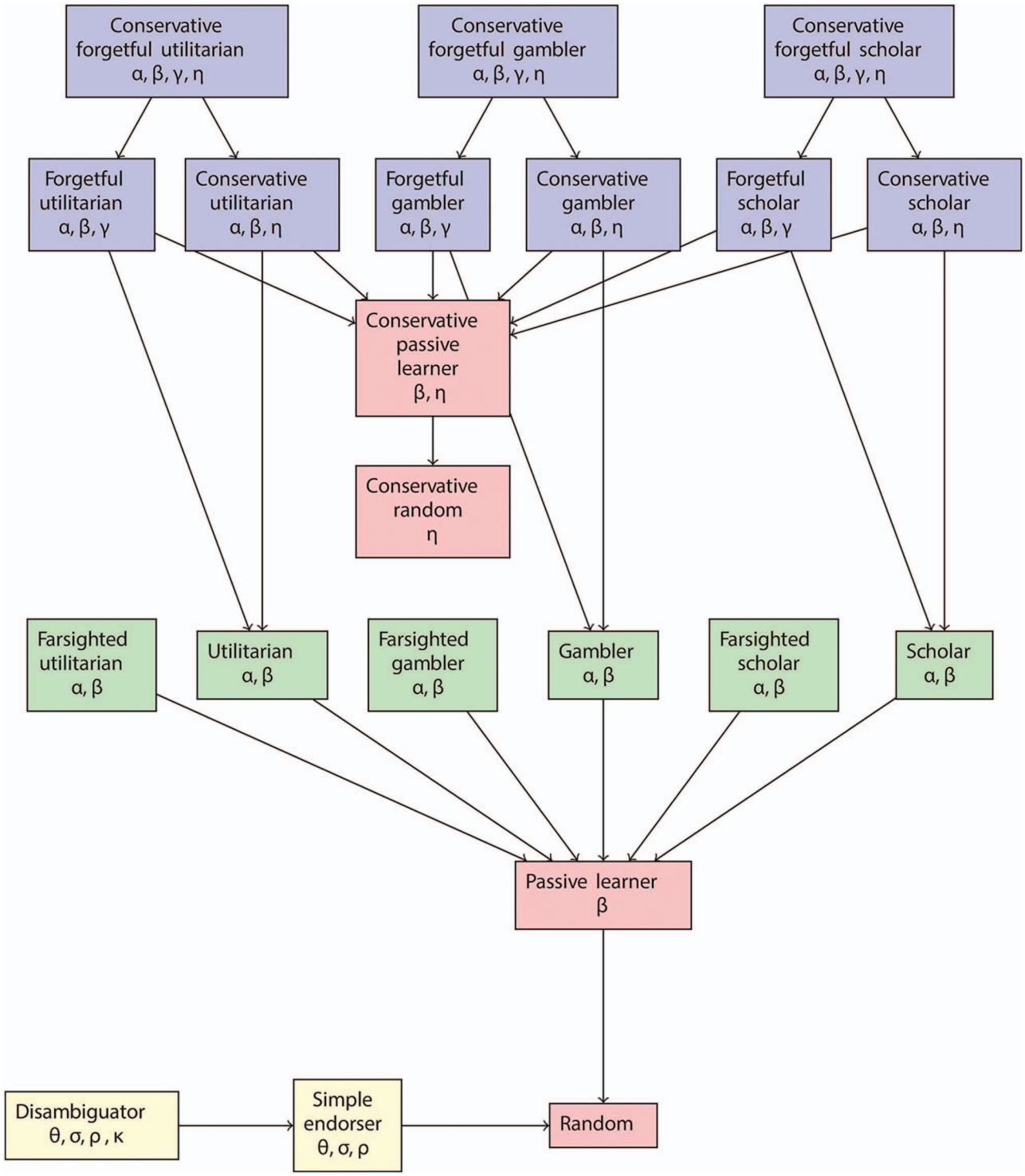

Figure 10. Schematic of the relationships between the models. Each model is nested within its parents and lists its fitted parameters. Blue rectangles (top two rows) indicate "bounded" models, green rectangles (fifth row) indicate "ideal" models, red rectangles (third, fourth, sixth, and seventh rows) indicate "null" models, and yellow rectangles (seventh row) indicate "heuristic" models. Arcs representing the nesting of the conservative null models in the nonconservative null models are omitted for clarity. See the online article for the color version of the figure.

utility gain U (Equation 8) in the utilitarian models, the probability gain Ψ (Equation 9) in the gambler models, and the information gain H (Equation 10) in the scholar models. Note that these quantities are computed from the distributions $p_t(g|q, o)$ and $p_t(g) = p_{t-1}(g|q_{t-1}, o_{t-1})$ as defined in Equation 11. We assume that chosen interventions are based upon these values through a variant of Luce's choice rule (Luce, 1959), such that the probability a learner selects intervention q at time t is given by

$$p(q_t = q) = \frac{e^{\alpha v_{qt}}}{\sum_{k=1}^{n} e^{\alpha v_{kt}}} \tag{12}$$

The parameter α controls how consistent the learner is in picking the intervention with the maximum expected value. As $\alpha \to \infty$ the probability that the learner picks the intervention with the highest expected value approaches 1 and the probability of picking any other intervention drops toward 0. If $\alpha = 0$, then the learner picks any intervention with an equal probability; that is, $p(q_t = q) = 1/n$, for all $q \in Q$.

Marking causal beliefs and conservatism. All expectancy-based models assume that learners' marked causal links are a noisy reflection of their current belief regarding the true causal models, as reflected by the posterior distribution $p_t(g|q_t, o_t)$. However, rather than using $p_t(g|q_t, o_t)$ directly, we allow for the possibility that the marking of causal beliefs may be subject to *conservatism*.

To allow for conservatism, we assume marked causal beliefs reflect a conservative probability distribution $p_t^*(g \mid q_t, o_t)$, which is a distorted version of the current distribution $p_t(g|q_t, o_t)$ in which the probability of causal structures consistent with the already marked causal links is relatively increased. Technically, this is implemented by multiplying the probability of consistent causal graphs by a factor $\eta \in [0, \infty]$ and then renormalizing the distribution.[8] The conservative probability distribution is given by:

$$p_t^*(g \mid q_t, o_t) = \frac{\eta^{I[g \in C]} p_t(g \mid q_t, o_t)}{\sum_{g' \in G} \eta^{I[g \in C]} p_t(g' \mid q_t, o_t)} \tag{13}$$

where $I[g \in C]$ is an indicator function with value 1 if the structure g is consistent with the currently marked links, and 0 otherwise. Marked links are assumed to be selected based on this conservative distribution. Then, this distribution is used to compute the values of the subsequent intervention options. For $\eta > 1$, sticking with already specified links is more likely than changing them all, other things being equal, while if $0 \leq \eta < 1$, this would lead to anticonservatism. Unlike forgetting, which has an effect that accumulates over trials, the conservative distortion is applied "temporarily" on each trial when marking beliefs and choosing the next intervention, but discarded thereafter, such that the prior on trial $t + 1$ is $p_{t+1}(g) = p_t(q|q_t, o_t)$ and not $p_t^*(q \mid q_t, o_t)$. By setting $\eta = 1$ we get a model that assumes participants are not neither conservative nor anticonservative.

As for interventions, we assume that a learner marks causal links through a variant of Luce's choice rule. The marking of links on each trial was optional, and initially all links were unspecified. As a result, links were often left unspecified, in which case a set of models S, rather than a single causal model, is consistent with the marked links.[9] To capture this, the models marginalize over all structures consistent with the links marked on a trial:

$$p(\text{Stated-belief}_t) = \frac{\sum_{g \in S} e^{\beta p_t^*(g \mid q_t, o_t)}}{\sum_{g' \in G} e^{\beta p_t^*(g' \mid q_t, o_t)}} \tag{14}$$

For example, if the participant has marked $A \to B$ but has so far left A-C and B-C unspecified, then the model sums over the probabilities of all the graphs that are consistent with this link. If a participant has not marked any links then their belief state for that time point trivially has a probability of 1.[10] By setting β to zero we get a model that assumes that participants are unable to identify causal links above chance regardless of what evidence they have seen.

Null, ideal and bounded expectancy models. In summary, the expectancy-based models have four free parameters: α controls the degree to which the learner maximizes over the intervention values, β controls the degree to which the learner maximizes over their posterior with their link selections at each time point, γ controls the extent to which participants discount or forget about past evidence and η controls the extent to which participants are conservative about the causal links they mark. See Figure 11 for a flowchart of how the full expectancy-based models work. By constraining the models such that combinations of these parameters are fixed, a nested set of expectancy-based models is obtained (see Figure 10). Fixing parameters to a priori sensible values can be important. For instance, we can assess whether a learner is forgetful by comparing a model in which the γ parameter is estimated to one in which the parameter is fixed to $\gamma = 0$.

A useful way to break down these models is divide them into null models, ideal models, and psychologically bounded models. We will call models with one or both of α and β fixed to zero null models. These models either assume that no active intervention selection takes place ($\alpha = 0$, interventions are selected randomly) and/or that no successful causal learning takes place ($\beta = 0$). We will call the models in which γ is fixed to 0 and η to 1 ideal models. These models are ideal in the sense that they set aside potential psychological constraints and so are at the computational level according to Marr's hierarchy (Marr, 1982). Comparing just

[8] This parameter works only once participants have registered their beliefs about at least some of the links, but this is the case on 91% of trials. On 76% of these trials participants had registered a belief for all three links, meaning that the conservativeness parameter upweights the subjective probability of this one structure while participants are selecting their new belief state and choosing the next intervention. This means that even if a learner's posterior is relatively flat due to forgetting, structures consistent with his (or her) marked links still stand out, leading him to behave as if he has selectively remembered information confirming these hypotheses. On the 24% of trials in which some but not all links remained unspecified, the conservativeness parameter led to the structures consistent with the established links being upweighted, leading the learner to favor interventions likely to distinguish between these options: Concretely, there would be 9 structures consistent with one specified link and 3 structures consistent with two specified links.

[9] A side effect of this aspect of the design is that we have more data on some participants than others. Those who rarely marked links before the end of the task reveal less information about how their belief at one time point influences their belief at subsequent time points.

[10] Cyclic Bayesian networks cannot be defined within the Bayesian network framework, and participants were instructed that they were impossible during the instructions. Therefore, on the 4.3% of trials in which participants marked a cyclic structure ([A → B, B → C, C → A] or [A → C, C → B, B → A]), their belief state was treated as unspecified so that the model did not return a likelihood of zero for that participant.

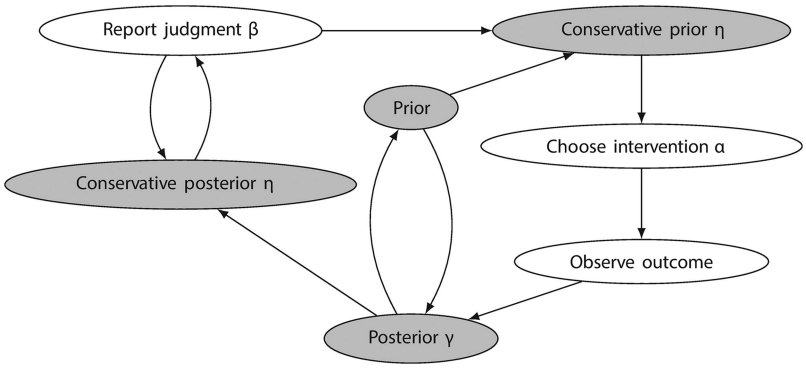

Figure 11. A flowchart of the expectancy based models. White nodes are observed quantities; gray nodes are unobserved quantities. Clockwise from the top right: The causal judgment reported at the previous time point and the prior distribution over causal structures combine to form a conservative prior. This is used to choose the next intervention. The outcome of the intervention is observed, and this is integrated with the (partially forgotten) prior to arrive at the posterior distribution over causal structures. The posterior and the previously reported judgment are mixed to form a conservative posterior that influences the new judgment. The posterior becomes the new prior, and then the process repeats.

these models addresses the question of which computational level problem participants' actions and judgments suggest they are (approximately) solving. Finally, we will call the full models in which one or both of γ and η are fit to the data bounded models. These models are bounded in the sense that they attempt to capture how psychological processing constraints potentially distort or change the computational problem, allowing us to explore how people might mitigate this in their intervention strategies.

Sensible evaluation of the bounded models requires different null models. For example, it may often be the case that someone is conservative about their beliefs despite those beliefs being completely random ($\beta = 0$). Alternatively people might be conservative *passive* learners yet unable to select sensible interventions, choosing interventions that are not more useful than chance ($\alpha = 0$). In these cases, we would have no reason to ascribe scholarly, gamblerly, or utilitarian behavior despite our models capturing some systematicity in participants' data.

Far-sighted scholars, gamblers and utilitarians. As mentioned in Greedy or Global Optimization? the values of different actions depend to some extent on how far the learner looks into the future. Computing expected values looking more than two steps ahead quickly becomes intractable even in the three-variable case, but we were able to compute the ideal two-step-ahead models for the three measures.[11] This allows us to check if there is evidence that people are able to look more than one step ahead when choosing interventions. Accordingly, we fit additional farsighted utilitarian, gambler, and scholar models in which the intervention values for looking one step ahead were replaced with those looking two steps ahead. We can compare these to the one-step-ahead ideal models to see if there is evidence that participants were planning more than one step ahead. We did not include freely estimated forgetting (γ) or conservatism (η) parameters in these models, because recomputing the two-step-ahead intervention values on the fly for different γ and η increments was prohibitively computationally expensive.

Heuristic Models

In addition to the various expectancy-based models described above, we explored whether people's intervention patterns can be well described by heuristic active learning models. By heuristic models, we mean models in which probabilities are not explicitly represented and values are not calculated for different interventions. Instead, these models assume that learners follow simple rules of thumb in order to choose their interventions, and update their causal models, without performing computationally demanding probabilistic information integration (Gigerenzer, Todd, & the ABC Research Group, 1999). Here we fit two models, the first nested in the second.

The simple endorser. One way to significantly simplify the causal learning problem is to ignore the dependencies between the causal connections in the possible graphs (Fernbach & Sloman, 2009). Thus, if intervention $A+$ is performed and both B and C activate this can be seen as evidence for an $A \rightarrow B$ connection and, independently, evidence for an $A \rightarrow C$ connection. In contrast a full Bayesian treatment would also raise the probability of other hypotheses (the an ABC and ACB chains and fully connected networks). Another way to simplify the problem is to ignore the Bayesian accumulation of probabilistic evidence and rather update belief directly to be consistent with the latest evidence. Concretely, in this task these assumptions would lead to people simply clamping variables on, one at a time, and adding links to any other nodes that activate as a result (removing any links to other nodes which do not activate as a result). We can operationalize this with a three parameter model (see Figure 12) which selects one of the simple interventions with probability $\theta \in [0, 1]$ or else selects anything else with probability $1 - \theta$. With a probability $\sigma \in [0, 1]$ the belief

[11] These expectancies are computed recursively, taking the maximum over the second set of interventions and passing these values back to the first set of interventions. See www.ucl.ac.uk/lagnado-lab/neil_bramley for code.

state is updated such that it becomes the prior belief state B plus links L from the current clamped node to any activated nodes (and minus those not in L but in B), while with probability $1 - \sigma$ it either: stays the same (with probability $\rho \in [0, 1]$) or takes any other state (with probability $1 - \rho$). A potential strength of this model in fitting the data is that it leads to systematic misattribution of a fully connected network when the true structure is a chain, a behavior exhibited by many participants. This happens because when the true structure is a chain, intervening on the root node will tend to lead to both other nodes activating, leading to the addition of direct links from the root node. When the middle node is intervened on this will tend to activate the last node, leading to the addition of the third link. To the extent that participants frequently act in this way, θ and σ will be high and the model fit will be good, and to the extent that they act in other ways the model will approach the fit of the null model in which beliefs and actions are selected at random.

The disambiguator. As we show earlier in the paper, controlling variables is a hallmark of scientific thinking, and a necessary part of successfully disambiguating causal structures (Cartwright, 1989; Kuhn & Dean, 2005). In this task this takes the form of a controlled intervention in which one node is clamped on and another clamped off (e.g., $A+$ $B-$), normally performed after observing some confounding evidence (i.e., when you clamp one node on and both other nodes activate). This action tests whether the node that remains unclamped is a direct effect of the node which is clamped on (see Figure 3), and thus *disambiguates* between the structures which could explain why both unclamped nodes activated on the previous trial. In the general case, the putative cause node would remain clamped on, a single putative effect node would be left unclamped and the other $N - 2$ nodes would be clamped off.

The model is operationalized as selecting $A+$, $B+$, or $C+$ or a disambiguation step (e.g., $A+$ $B-$) with probability θ and something else with probability $1 - \theta$ (see Figure 13). Propensity to select a simple endorsement step rather than a disambiguation step is governed by a fourth parameter $\kappa \in [0, 1]$. If a disambiguation step is performed and the unclamped node does not activate, then any connection from the activated node to the unclamped node is removed with probability σ. The belief update step is otherwise the same as for the simple endorser.

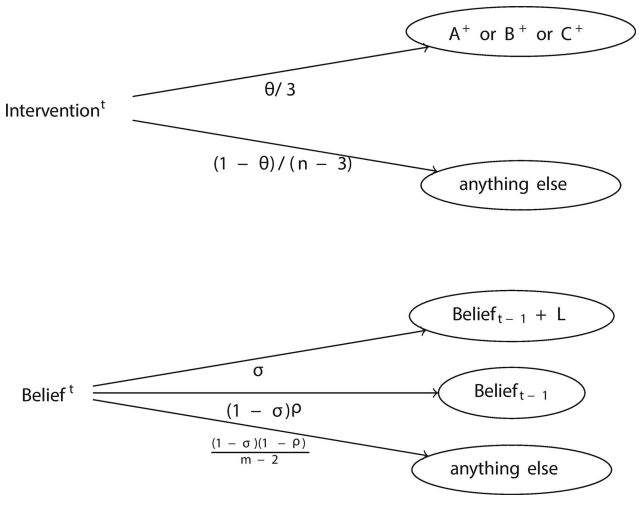

Figure 12. Process trees for the simple endorser. The learner follows an arrow with the probability written under the arrow and takes the action in the end node.

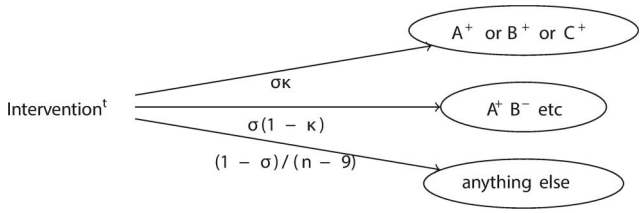

Figure 13. Process tree for the intervention selection step for the disambiguator. Belief update step is the same as for the simple endorser.

Model Estimation and Comparison

All models were fitted to individual's data by maximum likelihood estimation.[12] These consist of four nested sets, one for each of the three expectancy measures (utility gain, probability gain, and information gain) and one for heuristic models. Each nested model has between zero and four parameters.

McFadden's pseudo-R^2 is computed for each model to give an idea of its goodness of fit.[13] This measure does not penalize model complexity so models are compared throughout using the Bayesian information criterion (BIC, Schwarz, 1978) which can be used to compare both nested and nonnested models (Lewandowsky & Farrell, 2010).

Model Fit Results

Full results of the model fits are contained in Table 3. Overall, the best fitting model was the fully bounded scholar model based on maximizing information gain with both conservatism and forgetting (hereafter *CF scholar*). This model had a pseudo-R^2 of .47, indicating a very good fit to the data,[14] and was the best fitting model for 103 out of the 139 participants over Experiment 1, 2a, and 2b according to the BIC. Of the 36 participants that were not best described as CF scholars, 24 were in Experiment 1 and many of these were best fit by the *conservative random* null model. See Figure 14 for a visual comparison of the scholar model with either or both of forgetting and conservatism as fit to a participant in Experiment 2a. Looking at their average scores, we see that those best described as CF scholars perform much better than those who are not CF scholar ($M = 11.3$), non-CF scholar, $M = 6.6$, $t(137) = 7.1$, $p < .0001$.

[12] We used the Nelder–Mead algorithm to numerically maximize the likelihood, as implemented in R's optim function. Optimization was validated by repetition with different starting parameter values.

[13] McFadden's pseudo-$R^2 = 1 - \frac{\log L(M_{\text{full}})}{\log L(M_{\text{minimal}})}$, where $L(M)$ denotes the likelihood of model M. The minimal model M_{minimal} is random (no learning) in Table 3, where both actions and endorsements are completely random.

[14] Values between .2 and .4 are considered a good fit (Dobson, 2010).

Table 3
Total BICs, Median Parameter Estimates, and Pseudo-R^2 Values for All Models

Model	BIC	α	β	γ	log(η)	pseudo-R^2	Best fit	Best ideal	Best heuristic
Random	98,396					0	0	5	1
Passive	84,487		5			0.15	0	9	0
U	79,905	6.1	4.9			0.2	0	1	
G	80,718	11	4.9			0.19	0	1	
S	75,236	6	4.9			0.25	1	117	
Farsighted U	78,576	9.3	4.9			0.21	0	0	
Farsighted G	78,715	13	4.9			0.21	0	1	
Farsighted S	76,229	4.1	4.9			0.24	0	5	
C random	76,574				4.5	0.23	9	0	28
C passive	75,453		5.7		6.7	0.25	0	0	0
CCU	72,635	7.1	5.8		3.1	0.28	0		
CG	73,739	12	5.5		2.8	0.27	0		
CS	68,190	6.7	5.8		3.3	0.33	0		
FU	67,694	11	14	0.68		0.33	4		
FG	70,027	32	12	0.79		0.31	1		
FS	64,039	7.7	11	0.47		0.37	2		
CFU	58,413	15	16	0.93	2	0.43	5		
CFG	61,680	53	17	0.97	1.3	0.4	6		
CFS	54,757	8.3	13	0.81	2.3	0.47	103		
		θ	κ	σ	ρ				
Simple	64,985	0.85	0.22	0.63		0.36	5		61
Disambiguator	64,100	0.95	0.22	0.63	.96	0.37	3		49

Note. Letters in the first column indicate (C)onservative and/or (F)orgetful, (U)tilitarian, (G)ambler, and (S)cholar models. The Best fit column gives the number of participants best fit by each model according to the Bayesian information criterion (BIC). The Best ideal column gives the same statistics as the Best fit column but includes only the ideal learner models and appropriate null models; the Best heuristic column does the same for the heuristic models.

Inspection of the forgetful models suggests that participants forgot a large amount of the evidence they received with median forgetting rates (γs) of .68, .79 and .47 for the utilitarian, gambler, and scholar models respectively. When paired with conservatism in the conservative models, forgetting rates become even higher. This makes intuitive sense, because high conservatism can result in a high probability for already marked links that would otherwise have to be due to participants maintaining more of the true posterior (see Figure 14). Looking at the parameter estimates of the CF scholar model, more forgetful people were also more conservative, with a significant rank–order correlation between γ and η ($\rho = .43$, $p < .0001$). In addition both forgetting and conservatism are negatively correlated with participants' overall scores, $\rho = -.70$, $p < .0001$ for γ and $\rho = .53$, $p < .0001$ for η.

Looking across experiments, we see that median forgetting (γ) in the forgetful scholar model drops considerably going from .71 in Experiment 1 to .30 in Experiment 2a and slightly further again to .25 in Experiment 2b. Naively we might expect that participants in Experiment 2b should not need a forgetting parameter, because they could see all of their past actions and outcomes. However, only one participant in Experiment 2b and none in Experiment 2a or 1 were better fit by a model without a forgetting parameter, meaning that the parameter still did work even for participants in Experiment 2b.[15] Rather, we conclude that "forgetting" in our models does not just capture people's inability to recall past evidence. More generally, we think it captures a recency bias or tendency to attend disproportionately toward newer over older evidence regardless of whether the older evidence is still accessible.

Ideal models. Although the CF scholar model performed best overall, the scholar, gambler, and utilitarian model predictions were often relatively similar when all four parameters were included. This could be because for flatter posteriors, the intervention values according to these models do not differ as much as they do when the posteriors are more peaked. Comparing predictions of the models with increasing forgetting rates confirms this (see Figure 15), with the level of agreement about the best intervention(s) and the average bivariate correlation between values of the different interventions all approaching 1 as forgetting rate increases toward 1. For a clearer assessment whether learners are best described as scholars, gamblers or utilitarians, we turn to a comparison of the ideal versions of these models (without forgetting or conservatism).

Considering only the ideal models and the relevant null models (see Figure 10 and the Best ideal column in Table 3), the scholar model clearly outperforms the utilitarian and gambler models. In this set of models, the scholar best captures 117 out of the 139 participants, including almost all high-scoring participants (scholar $M = 10.8$, nonscholar $M = 6.4$). Nine of the poorest participants (average score 5.5) were also better described as achieving some learning, despite failing to select interventions more useful than chance (passive learner), but none were best fit by the completely chance-level random model, in which both α and β were set to zero.

Looking across experiments, we see that median αs for the ideal scholar model, controlling maximization over intervention values, increase from 5.2 ($SD = 2.7$) in Experiment 1 to 6.6 ($SD = 3.0$) in Experiment 2a and 7.1 ($SD = 3.0$) in Experiment 2b. Likewise,

[15] This participant identified every connection correctly and was best described as an ideal scholar.

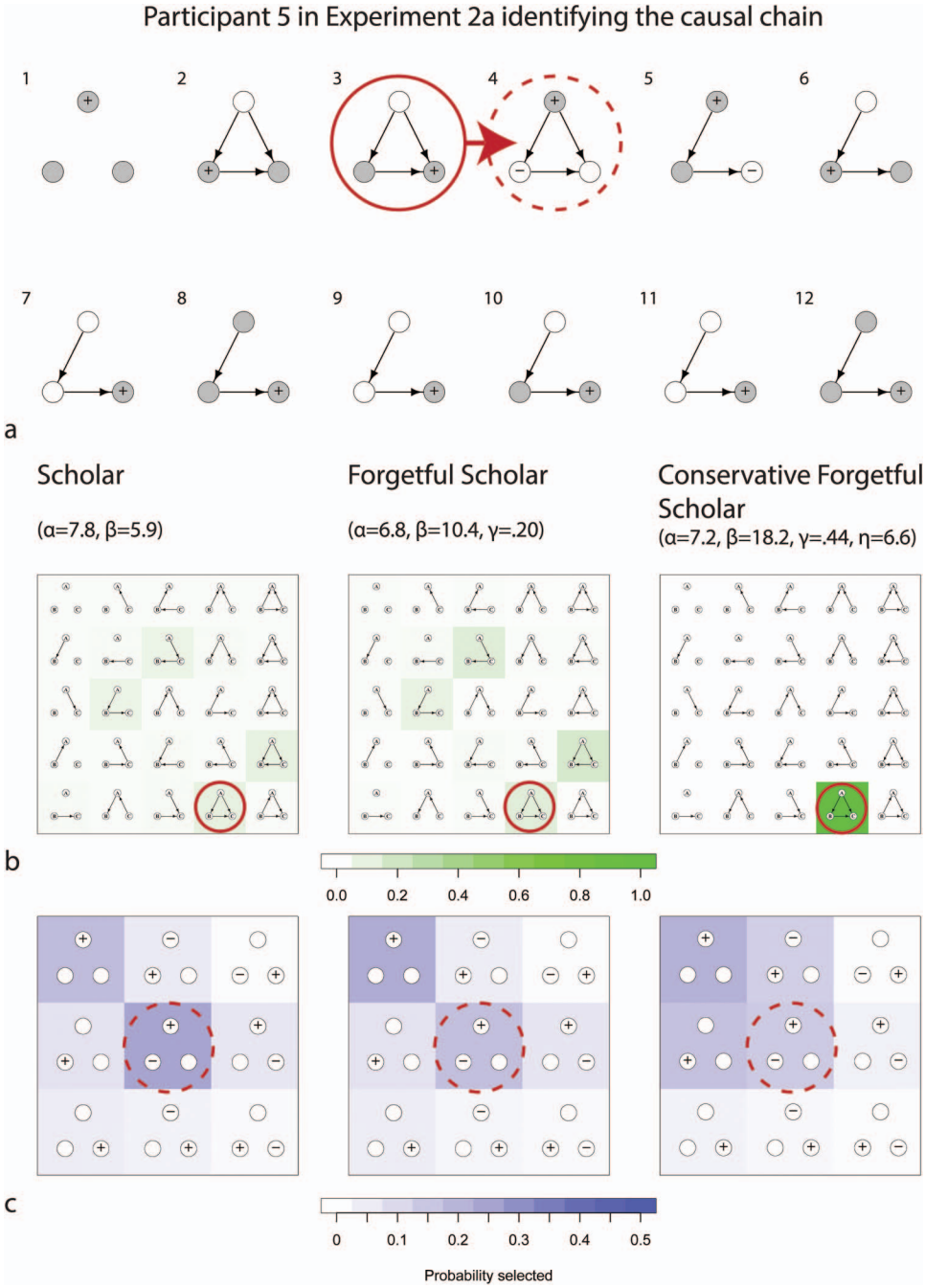

Figure 14. Visual comparison of fitted models. a. Participant 5 in Experiment 2a, identifying the chain ($A \rightarrow B$, $C \rightarrow B$) structure. Plus and minus symbols indicate interventions, gray nodes indicate the resultant activations, and the arrows replicate those marked by the participant at each time point. b. The probability that the participant registers each causal structure according to the scholar, forgetful scholar, and conservative forgetful scholar models (their actual choice is the full circle). c. The probability of selecting each of the simple and controlled interventions on the next test (actual choice is the dashed circle). See the online article for the color version of the figure.

median βs, controlling maximization over the posterior under the scholar model, increase from 3.5 ($SD = 26$) in Experiment 1 to 5.9 ($SD = 3.5$) in Experiment 2a and 6.5 ($SD = 2.9$) in Experiment 2b. This suggests that when the task was clarified and especially when summary information was provided, participants' interventions judgments were closer to those arising from expected information maximization and Bayesian inference.

There is no evidence that people were able to look more than one step ahead in this task though, with across-the-board worse fits for the farsighted scholar, gambler, and utilitarian models and only

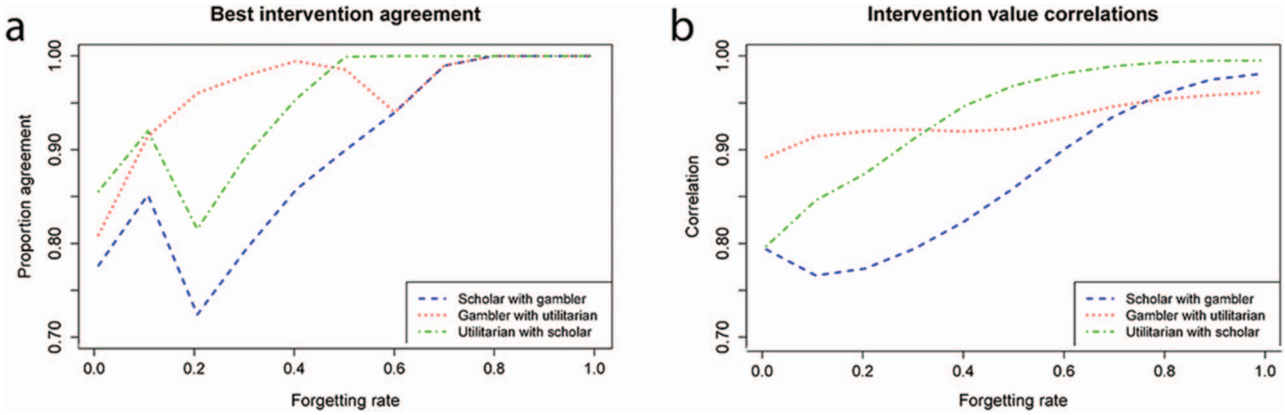

Figure 15. a. Proportion of participants' tests on which the forgetful scholar, gambler, and utilitarian models agree about the most useful intervention as a function of models' forgetting rate. b. Mean correlation between intervention values according to these models for all participants' tests, as a function of forgetting. See the online article for the color version of the figure.

6/139 participants best fit by one of these models rather than the one-step-ahead or null models. These were not the most gifted participants, scoring an average of only 7.5, suggesting that the resemblance between their interventions and to those favored by the two-step-ahead expectancies was accidental.

In summary, comparing the ideal learner models shows that successful causal learners' actions and causal judgments are more closely related to the computational level problem of reducing uncertainty than those of maximizing probability or utility gain.

Heuristic models. When comparing the full set of models, few of the participants were best described by either of the heuristic models. Nevertheless, these models fit relatively well despite their algorithmic simplicity, with BIC values in the range of the forgetful Bayesian models. Ignoring the expectancy-based models, we can, similarly as previously for the ideal models, compare the heuristic models against the relevant null models (see Table 3, last column). From this we can see that the better fitting heuristic overall is the *disambiguator*. However, more individual participants are better described as simple endorsers (61) than disambiguators (49), with the remainder being described by the conservative random null model. The majority (18/28) of those better described as conservative random are in Experiment 1 and had average scores of only 6.4. Over Experiments 1, 2a, and 2b, those described as disambiguators do slightly better than those described as simple endorsers, $t(101.6) = -2.0$, $p = .04$. Disambiguators used complex interventions on 8.8% of trials (14.8% for chain and fully connected models) while simple endorsers rarely or never used complex interventions (1.3% of the time; 2% of the time for the chain and fully connected models).

General Discussion

Overall, our analyses suggest that the majority of people are highly capable active causal learners, both in terms of selecting useful interventions and in terms of learning from them. Having identified task confusion as the cause for many of the poorer performances in Experiment 1, we found that with an in-task reminder of the instructions in Experiment 2, almost all participants performed very well. Allowing participants to see the results of their past tests did not make a significant difference to performance but did significantly reduce task completion time. because performance was already near ceiling in Experiment 2, we can see the quicker completion times suggesting that the history of past trials did make the task somewhat less demanding.

Simulations of efficient utility-, probability-, and information-maximizing active learning showed that starting with simple interventions and gradually switching to more focused controlled interventions made for an efficient interventional strategy. We see participants exhibiting this same pattern, starting with almost exclusively simple interventions and gradually using more controlled ones as they narrow down the space of possible structures. Participants' interventions were also somewhat sensitive to the structure being learned, with more controlled interventions being selected on the chain and fully connected networks, where it was very hard to identify the correct causal structure without at least one controlled intervention. While participants were generally less inclined to select controlled interventions than the benchmark models, this is consistent with their learning being slower and more imperfect, as is reflected by our fitted models.

We can think of simple interventions as open-ended tests. They do not test any one hypothesis in particular and have multiple possible outcomes, each of which can be consistent with several different causal interpretations (e.g., if there are two activations following a simple intervention, these could result from a chain, common effect, or fully connected structure). However, simple interventions are powerful at first because they quickly reduce the space of likely models. In contrast, controlled interventions can be seen as more focused tests. They have only two possible outcomes and lend themselves to distinguishing unambiguously between two or three causal structures that perhaps differ by only one causal connection. This progression from open-ended to more focused testing gels with a picture of people as natural scientists, first exploring the space and identifying a candidate causal model, then progressively refining this with focused experiments. We found that the propensity to select controlled tests was closely linked to high performance, suggesting that only more sophisticated causal learners would progress from the exploratory stage to the stage of

performing specific controlled hypothesis tests. The idea that controlled interventions are more cognitively demanding than simple ones is supported by research on complex control (e.g., Osman, 2011), where ability to recognize that one must simultaneously manipulate two variables to control a system is difficult for many people.

We found that participants had a strong tendency to mistake chains for fully connected structures across both experiments. One reductive explanation for this is that some participants may have misunderstood the task demands, interpreting links as meaning that the parent node is a *direct or indirect* cause of the child node. However, the instructions were clear on this point, demonstrating the way in which clamping an in-between node off would block activation passing along a chain. Instead, we conclude that this mistake is a marker for many participants' heuristic causal learning strategies. This is confirmed by the large number of participants whose actions and judgments are better described by the simple endorsement heuristic that systematically overconnects chains rather than the disambiguation heuristic.

We compared computational models of efficient causal learning, driven by three plausible measures of intervention values: expected information gain, probability gain, and utility gain. Overall, the models driven by information gain (scholars) better fit the large majority of participants' interventions than models driven by probability gain (gamblers) or utility gain (utilitarians). This was particularly clear looking at the ideal models (without forgetting or conservatism). This means that, however participants were choosing their interventions and updating their beliefs, they were managing to do so in a way which broadly approximated the solution to the computational level problem of maximizing information rather than that of minimizing error or maximizing utility.

Venturing one rung down the ladder from the computational level toward psychological process (Jones & Love, 2011; Marr, 1982), we explored "bounded" versions of our models. We included forgetting and conservatism parameters capturing the idea that people might be biased in their learning by plausible memory and processing constraints. The fit of our models was greatly improved by inclusion of these parameters and including both parameters led to much better overall fits than including only one. The two parameters were correlated, supporting the idea that they complemented one another: for example, the more forgetful a learner is about past evidence, the more conservative they need to be in their beliefs in order to be an effective learner. Therefore, these models provide an account of how moderate forgetting of old evidence paired with appropriate conservatism about existing causal beliefs can lead to effective active causal learning.

Allowing participants to draw and update models as they went along may have affected their learning, perhaps distracting them, or leading them to place more emphasis on earlier marked links. Furthermore, while we accept that the beliefs reported by participants at each time point are at best noisy markers of their actual beliefs about the true structure, we feel that these are largely unavoidable aspects of tracing beliefs throughout learning. We tried to minimize the extent to which eliciting beliefs distracted participants by making the step optional and hoping that participants would voluntarily record their beliefs as an aid to memory. It seems this was what most participants did, as links were drawn on 91% of tests, and neither varied wildly nor remained static from trial to trial. As a result we have been able to explore patterns of sequential causal learning in an unprecedented level of detail.

Taking another step toward the process level, we also looked at whether participants' actions could be reasonably captured by simple heuristics. We noted that simple endorsement (Fernbach & Sloman, 2009), based on local computation, could capture much of the behavior of many participants. This may explain why so many participants judged the fully connected structure when the true structure was the chain. However, some participants also performed the crucial controlling disambiguation steps, which cannot be easily captured in a local computation framework. We operationalized this here as an alternative step occasionally performed at random. However, we note that a disambiguator type model has the potential to be refined by incorporating sequential dependence. For instance, a natural hypothesis is that disambiguation steps are most likely to be performed following ambiguous evidence (i.e., multiple activations). Another possibility is that learners are likely to perform disambiguation steps with the same node clamped on as they had clamped on for the step that generated the ambiguity. However, further refining the heuristic models in the current context is likely to make them increasingly indistinguishable from our expectancy-based models. To confidently identify people's heuristic strategies we will need to look at learning problems with a larger number of variables and the potential for larger divergences between heuristics and computational level models.

With these experiments and analyses we have begun the process of studying active causal learning behavior, starting with a simple open-ended experiment (Experiment 1) and more controlled follow-up (Experiment 2). Having motivated and constructed models of participants' actions and judgments at computational and process levels, the next steps will be to come up with controlled tests that allow us to rigorously test some of these predictions. For example, an avenue of future work will be to look at the range of environments within which heuristic strategies are effective. We hypothesize that the extent to which one must disambiguate (or control for other variables) depends on how noisy, complex or densely causally connected the environment is. For more than around 5 or 6 variables, explicit calculation of expectancies becomes intractable while the calculations required by the active causal learning heuristics remain computationally trivial. In everyday life people have to deal with causal systems with many variables, far more than would plausibly allow explicit expectancies to be computed. One way people might achieve this is by performing an appropriate mixture of these "connecting" simple endorsement and "pruning" disambiguation steps.

Conclusions

In this paper we asked how people learn about causal structure through sequences of interventions. We found that many participants were highly effective active causal learners, able to select informative interventions from a large range of options and use these to improve their causal models incrementally over multiple learning instances. We found that successful learners were able to make effective use of controlling double interventions as well as simple single interventions, doing so increasingly as they narrowed down the hypothesis space. The large majority of participants acted like scholars, choosing interventions likely to reduce their uncertainty about the true causal structure, rather than to increase

their expected utility or probability of being correct. We found no evidence that people were able to plan ahead when choosing interventions. We formulated bounded models that include forgetfulness and conservatism. These show that people exhibit recency when integrating evidence but also suggest that they mitigate this to a large extent by being appropriately conservative, preferring causal structures consistent with previously stated beliefs. Finally, we identified simple endorsement and disambiguation as candidate components of heuristics for active causal learning.

References

Baddeley, A. (1992, January 31). Working memory. *Science, 255,* 556–559. doi:10.1126/science.1736359

Baron, J. (2005). *Rationality and intelligence*. New York, NY: Cambridge University Press.

Baron, J., Beattie, J., & Hershey, J. C. (1988). Heuristics and biases in diagnostic reasoning: II. Congruence, information, and certainty. *Organizational Behavior and Human Decision Processes, 42,* 88–110. doi:10.1016/0749-5978(88)90021-0

Bruner, J. S., Jolly, A., & Sylva, K. (Eds.). (1976). *Play: Its role in development and evolution*. New York, NY: Basic Books.

Cartwright, N. (1989). *Nature's capacities and their measurement*. New York, NY: Oxford University Press.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review, 104,* 367–405. doi:10.1037/0033-295X.104.2.367

Coenen, A., Rehder, B., & Gureckis, T. (2014). Decisions to intervene on causal systems are adaptively selected. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 343–348). Austin, TX: Cognitive Science Society.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24,* 87–114. doi:10.1017/S0140525X01003922

Dobson, A. J. (2010). *An introduction to generalized linear models* (3rd ed.). New York, NY: Chapman & Hall.

Eberhardt, F., Glymour, C., & Scheines, R. (2005). On the number of experiments sufficient and in the worst case necessary to identify all causal relations among N variables. In F. Bacchus & T. Jaakkola (Eds), *Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence* (pp. 177–184). Arlington, VA: Association for Uncertainty in Artificial Intelligence Press.

Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmutz (Ed.), *Formal representation of human judgment* (pp. 17–52). New York, NY: Wiley.

Fernbach, P. M., & Sloman, S. A. (2009). Causal learning with local computations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 678–693. doi:10.1037/a0014928

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

Good, I. J. (1950). *Probability and the weighing of evidence*. London, United Kingdom: Griffin.

Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological Review, 116,* 661–716. doi:10.1037/a0017201

Gureckis, T., & Markant, D. (2009). Active learning strategies in a spatial concept learning game. In N. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 3145–3150). Austin, TX: Cognitive Science Society.

Gureckis, T., & Markant, D. (2012). Self-directed learning: A cognitive and computational perspective. *Perspectives on Psychological Science, 7,* 464–481. doi:10.1177/1745691612454304

Harman, G. (1986). *Change in view: Principles of reasoning*. Cambridge, MA: MIT Press.

Hartigan, J. A., & Hartigan, P. (1985). The dip test of unimodality. *Annals of Statistics, 13,* 70–84. doi:10.1214/aos/1176346577

Heckerman, D. (1998). A tutorial on learning with Bayesian networks. In M. Jordan (Ed.), *Learning in graphical models* (pp. 301–354). Cambridge, MA: MIT Press.

Holyoak, K. J., & Cheng, P. W. (2011). Causal learning and inference as a rational process: The new synthesis. *Annual Review of Psychology, 62,* 135–163. doi:10.1146/annurev.psych.121208.131634

Hyafil, L., & Rivest, R. L. (1976). Constructing optimal binary decision trees is NP-complete. *Information Processing Letters, 5,* 15–17. doi:10.1016/0020-0190(76)90095-8

Jones, M., & Love, B. C. (2011). Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences, 34,* 169–188. doi:10.1017/S0140525X10003134

Krynski, T. R., & Tenenbaum, J. B. (2007). The role of causality in judgment under uncertainty. *Journal of Experimental Psychology: General, 136,* 430–450. doi:10.1037/0096-3445.136.3.430

Kuhn, D., & Dean, D. (2005). Is developing scientific thinking all about learning to control variables? *Psychological Science, 16,* 866–870. doi:10.1111/j.1467-9280.2005.01628.x

Lagnado, D. A., & Sloman, S. (2002). Learning causal structure. In W. Gray & C. D. Schunn (Eds.), *Proceedings of the 24th Annual Conference of the Cognitive Science Society* (pp. 560–565). Mahwah, NJ: Erlbaum.

Lagnado, D. A., & Sloman, S. (2004). The advantage of timely intervention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 856–876. doi:10.1037/0278-7393.30.4.856

Lagnado, D. A., & Sloman, S. A. (2006). Time as a guide to cause. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 451–460. doi:10.1037/0278-7393.32.3.451

Lewandowsky, S., & Farrell, S. (2010). *Computational modeling in cognition: Principles and practice*. Thousand Oaks, CA: Sage.

Lindley, D. V. (1956). On a measure of the information provided by an experiment. *Annals of Mathematical Statistics, 27,* 986–1005. doi:10.1214/aoms/1177728069

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

Markant, D., & Gureckis, T. (2010). Category learning through active sampling. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 248–253). Austin, TX: Cognitive Science Society.

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

Meder, B., & Nelson, J. D. (2012). Information search with situation-specific reward functions. *Judgment and Decision Making, 7,* 119–148.

Meier, K. M., & Blair, M. R. (2013). Waiting and weighting: Information sampling is a balance between efficiency and error-reduction. *Cognition, 126,* 319–325. doi:10.1016/j.cognition.2012.09.014

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81–97. doi:10.1037/h0043158

Nelson, J. D. (2005). Finding useful questions: On Bayesian diagnosticity, probability, impact, and information gain. *Psychological Review, 112,* 979–999. doi:10.1037/0033-295X.112.4.979

Nelson, J. D., McKenzie, C. R., Cottrell, G. W., & Sejnowski, T. J. (2010). Experience matters information acquisition optimizes probability gain. *Psychological Science, 21,* 960–969. doi:10.1177/0956797610372637

Osman, M. (2011). *Controlling uncertainty: Decision making and learning in complex worlds*. Chichester, United Kingdom: Wiley-Blackwell.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. San Francisco, CA: Kaufmann.

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

Piaget, J. (2001). *The child's conception of physical causality* (J. Valsiner, Trans.). New Brunswick, NJ: Transaction. (Original work published 1930)

Puterman, M. L. (2009). *Markov decision processes: Discrete stochastic dynamic programming*. Hoboken, NJ: Wiley.

Schulz, L. (2001, October). *"Do-calculus": Adults and preschoolers infer causal structure from patterns of outcomes following interventions*. Paper presented at the meeting of the Cognitive Development Society, Virginia Beach, VA.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6,* 461–464. doi:10.1214/aos/1176344136

Shanks, D. R. (1995). Is human learning rational? *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 48*(A), 257–279. doi:10.1080/14640749508401390

Shannon, C. E. (2001). A mathematical theory of communication. *ACM Sigmobile Mobile Computing and Communications Review, 5,* 3–55. doi:10.1145/584091.584093

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

Sobel, D. M. (2003). *Watch it, do it, or watch it done: The relationship between observation, intervention, and observation of intervention in causal structure learning*. Manuscript submitted for publication, Brown University.

Sobel, D. M., & Kushnir, T. (2006). The importance of decision making in causal learning from interventions. *Memory & Cognition, 34,* 411–419. doi:10.3758/BF03193418

Sprites, P., Glymour, C., & Scheines, R. (1993). *Causation, prediction, and search*. Cambridge, MA: MIT Press.

Steyvers, M., Tenenbaum, J. B., Wagenmakers, E., & Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive Science, 27,* 453–489. doi:10.1207/s15516709cog2703_6

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011, March 11). How to grow a mind: Statistics, structure, and abstraction. *Science, 331,* 1279–1285. doi:10.1126/science.1192788

Waldmann, M. (2000). Competition among causes but not effects in predictive and diagnostic learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 53–76. doi:10.1037/0278-7393.26.1.53

Waldmann, M., & Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: Asymmetries in cue competition. *Journal of Experimental Psychology: General, 121,* 222–236. doi:10.1037/0096-3445.121.2.222

Wixted, J. T. (2004). The psychology and neuroscience of forgetting. *Annual Review of Psychology, 55,* 235–269. doi:10.1146/annurev.psych.55.090902.141555

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. New York, NY: Oxford University Press.

Received March 13, 2014
Revision received July 4, 2014
Accepted July 14, 2014 ■