Modelling active causal learning

Neil Bramley

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Master of Research
of the
University of London.

Department of Computer Science
University College London

August 16, 2013

Abstract

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 incentivised to learn the structure of probabilistic causal systems through free selection of multiple interventions. We develop models of participants' intervention choices and online structure judgements, using several measures of interventional utility and introducing plausible memory and processing constraints. We find that participants are best described by a model that acts to maximise information (rather than expected score, or probability of being correct); that forgets much of the evidence recieved in earlier trials; but that mitigates this by being conservative, preferring structures consistent with its earlier stated beliefs. We explore two heuristics that together explain how participants might be approximating this model without explicitly representing the hypothesis space.

Acknowledgements

Thanks to the UK PhD Centre in Financial Computing for funding this research. And, thanks to supervisors Dr David Lagnado, and Professor Peter Dayan, and collaborators Dr Maarten Speekenbrink and Dr Jonathan Nelson for helpful comments and support throughout.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

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

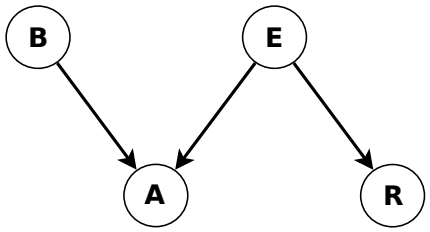
The idea that self direction is crucial to learning is well established in education and frequently noted in developmental psychology, where self-directed 'play' is seen as vital to healthy development (e.g. Bruner et al., 1976; Piaget and Valsiner, 1930). Additionally, recent work in cognitive science has shown that people learn categories and spatial concepts quicker by actively selecting informative samples (e.g. Gureckis and Markant, 2009, 1; Markant and Gureckis, 2010). However, it is in identifying causal structure that manipulating ones' data stream becomes really essential (Pearl, 2000; Woodward, 2003). To motivate this claim, consider observing a correlation between variables of interest, A and B. The correlation tells you that the variables are likely to be causally related but does not tell you in what way. Perhaps A causes B; perhaps B causes A; or perhaps they share some other common cause C. For many every day inferences there are strong spatial or temporal cues to causal structure, but in other cases these cues can be noisy, uninformative or unavailable. Many systems propagate too fast to permit observation of time ordering of component activations

(e.g. electrical systems); or have hidden mechanisms (e.g. biological systems, psychological processes); or noisy/delayed presentation of variable values (e.g. variations in gestation periods makes presentation time an unreliable guide to the transmission of a virus around a population, Lagnado and Sloman, 2006); while 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 and Tenenbaum, 2009). In the absence of such established cues, the direction of a causal connection cannot generally be established by mere observation but only 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). If manipulating A changes B then this is evidence that A causes B. Thus, ability to intervene provides active learners with a privileged route for obtaining useful causal information, one which is typically unavailable to machine learning research and which has therefore relatively little attention to date.

Several studies have found that people benefit from the ability to perform interventions (Lagnado and Sloman, 2002, 0, 0; Schulz, 2001; Sobel, 2003; Steyvers et al., 2003). However, only Steyvers et al's study explores how people select which interventions to perform, and only in the case of choosing of a single one-variable intervention after a period of observation. But, much real world causal learning is incremental, taking place gradually over many instances, making the dynamics of multi-shot interventional learning an important and understudied area. Additionally, single-variable interventions are not sufficient to learn all causal structures. Here we present an experiment in which participants perform multiple interventions and can manipulate as many variables as they wish with each intervention. Our design allows us to fit models to peoples' intervention choices and structure judgements, capturing how peoples' learning actions interact with their evolving causal beliefs over multiple learning trials. The experiment and analyses we present here are therefore significant extensions to the existing research along several dimensions. In order to motivate these extensions, we first briefly introduce the mathematical machinery of causal inference and intervention.

1.1 Causal Bayesian networks

Bayesian networks (Barber, 2012; Pearl, 2000; Sprites et al., 1993) provide a useful way to think about and quantify the role of causal information in inference. They are a natural extension of the probabilistic models approach (e.g. Chater and Oaksford, 2008) to the task of building structured representations. A Bayesian network is a parametrised directed acyclic graph (Figure 1.1). The nodes in the graph represent some variables of interest and the links between the nodes represent dependencies. Bayesian networks are defined by the Markov condition, which states: *each node is independent of all of its non-descendants 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.



$$p(A|B,E)p(R|E)p(E)p(R) \tag{1.1}$$

$p(A=1)$	B	E
.91	1	1
.82	1	0
.55	0	1
.1	0	0

$p(R=1)$	E
.999	1
.02	0

$p(B=1)$
.01

$p(E=1)$
.000001

Figure 1.1: The directed acyclic graph of a causal belief network capturing the causal relationships between a (B)urglar, (E)arthquake, burglar (A)larm, and a (R)adio report all of which $\in [1 = Present, 0 = Absent]$. Arbitrary probability tables are given for the variables. Adapted from Barber (2012, p34).

To capture causal structure, we interpret the directed links as causal connections going from cause to effect. Each node is parametrised with a table of probabilities, one for each state of the node given each state of each of its putative causes. Thus, if you

know the state of all of the causes and effects of a variable then you can ignore the states of any other variables in the graph as they will not provide any more information. For example, Figure 1.1 shows a causal belief network capturing relationships which could exist between four binary variables: a (B)urglar or an (E)arthquake might cause your burglar (A)larm to go off. However, an (E)arthquake will generally also cause a news bulletin to be put out on the (R)adio. If you know that there has been an earthquake and no burglary, then you can read off the probability of your alarm going off without worrying about whether there has been a radio broadcast. The probability of the state of any variable in the network given any set of knowledge about the other variables can be calculated using Bayes' theorem, summing over any unknown variables and conditioning on any known variables, ignoring any variables that are not causally relevant. Without causal structure, one would have to sum over all unknowns in the full joint distribution for every inference. Summing over everything in a model is of $O(2^N)$ complexity so is intractable for all but the most trivial problems. Thus, causal Bayesian networks formalise the intuition that organising ones' beliefs to reflect causal structure reduces the complexity of inference.

For example, suppose you come home and your burglar Alarm is ringing $(A = 1)$, you can calculate the probability that there was a burglar using,

$$p(y|x) = \frac{p(x, y)}{p(y)} \tag{1.2}$$

simply replacing $p(x, y)$ for the joint probability of the alarm and the burglar, and normalising by the marginal probability of the alarm:

$$\begin{aligned} p(B = 1|A = 1) &= \frac{\sum_{E,R} p(A = 1|B = 1, E)p(R|E)p(E)p(B = 1)}{\sum_{B,E,R} p(A = 1|B, E)p(R|E)p(E)p(B)} \\ &\approx .9 \end{aligned} \tag{1.3}$$

It is useful to think about the strength of a causal link as the probability that the cause will produce its effect in the absence of other causes (Cheng, 1997). This is called its *causal power*:

$$power_{c \to e} = \frac{p(e = 1|c = 1) - p(e = 1|c = 0)}{1 - p(e = 1|c = 0)} \tag{1.4}$$

where $e = 1$ denotes the presence of the effect, $e = 0$ denotes the absence of the effect, and $c = 1/c = 0$ denote the presence/absence of the cause.

This formalisation also makes it easy to think about how multiple independent causes might combine to produce their effects (Cheng, 1997; Griffiths and Tenenbaum, 2009). If we know the power of several causes of a single effect, and we believe these causes have independent chances to produce the effect, then the probability of the effect can be modelled as a noisy-OR function of these causes (Pearl, 1988)[1].

$$p(e = 1|c_1 = 1, c_2 = 1, ..., c_n = 1) = 1 - (1 - power_{c1})(1 - power_{c2})...(1 - power_{cn}) \tag{1.5}$$

Using the example above, the alarm has a *base-rate* probability of activation $p(e = 1|c = 0)$ of .1 (i.e. it can be caused by factors outside the scope of the current Bayesian network). Thus, we see that the power of the (B)urglar to cause the (A)larm to go off is $(.82 - .1)/(1 - .1) = .8$ and the power of the (E)arthquake to cause the alarm to go off is $(.55 - .1)/(1 - .1) = .5$. The probability of the alarm going off when there is both a burglary and an earthquake accordingly is $1 - (1 - .8)(1 - .5)(1 - .1) = .91$.

Learning causal structure can sometimes be achieved purely by observation. Given some data D, one can search over the space of possible causal structures G for the most probable ones (integrating over parametrisations θ). The posterior probability of some graph $g_i \in G$ is given by:

$$p(g_i|D) = \frac{\int_\theta p(D|g_i, \theta)p(\theta|g_i)p(g_i)d\theta}{\sum_{j \in G} \int_\theta p(D|g_j, \theta)p(\theta|g_j)p(g_j)d\theta} \tag{1.6}$$

The number of possible causal graphs rapidly becomes very large (25 3-variable, 543 4-variable, 29281 5-variable, etc). Efficient algorithms such as *network scoring* (Heckerman et al., 1995) and the constraint-based *PC algorithm* (Pearl, 2000; Sprites et al., 1993), sidestep some of the computational difficulties searching this space. However, without knowledge of the causal power of potential causal links, many structures are still observationally indistinguishable, or *Markov equivalent* (Pearl, 2000), meaning that observational learning methods equivocate over multiple structures equally

[1] Causal Bayesian networks can equally capture other combination functions (e.g. AND, exclusive-OR etc). In fact any arbitrary set of contingencies can be represented by setting individual table entries.

consistent with the same data. For example, $A \rightarrow B \rightarrow C$ and $A \leftarrow B \rightarrow C$ are both viable ways of structuring a model of a probability distribution in which A is independent of C given the value of B. This is not the case if, as in our experiment, the learner is assumed to know the causal power of variables. However, even without Markov equivalence, a large amount of data, exponential in the number of variables, is generally required to reliably identify causal structure from observation (Griffiths and Tenenbaum, 2009) and research suggests people struggle to learn causal structure from purely statistical information (e.g. Lagnado and Sloman, 2002). These considerations suggest that we must look beyond mere observation to explain peoples' remarkable ability to infer causal structure of their environment.

1.1.1 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. Interventions are modelled as *graph surgery* (Pearl, 2000, Figure 1.2). Any links going to an intervened on variable are temporarily severed, or removed. Then, the variable is set, or *clamped*, to a particular value. The resulting values of the unclamped variables are then observed and inference is performed as normal. To distinguish certain structures, one must clamp multiple variables in a single intervention, as we do when controlling for potential confounds in an experiment (Figure 1.3). Eberhardt et al. (2012) show that at most $log_2(N) + 1$ idealised experiments are sufficient to identify any causal structure connecting N variables, 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) that it can reliably rule various structures out. However, in this paper we are interested in *token*, single-sample experiments. This is because interventions achievable by real-world learners are prototypically single-sample affairs - e.g. pulling a lever, adding an ingredient, restarting the wifi-router, coming off a medication, taking the batteries out, etc. Therefore, where causal links are a bit noisy, repetitions of interventions will

be required to identify the correct causal structure with a high probability (although at most $log_2(N) + 1$ different *types* of intervention will be sufficient).

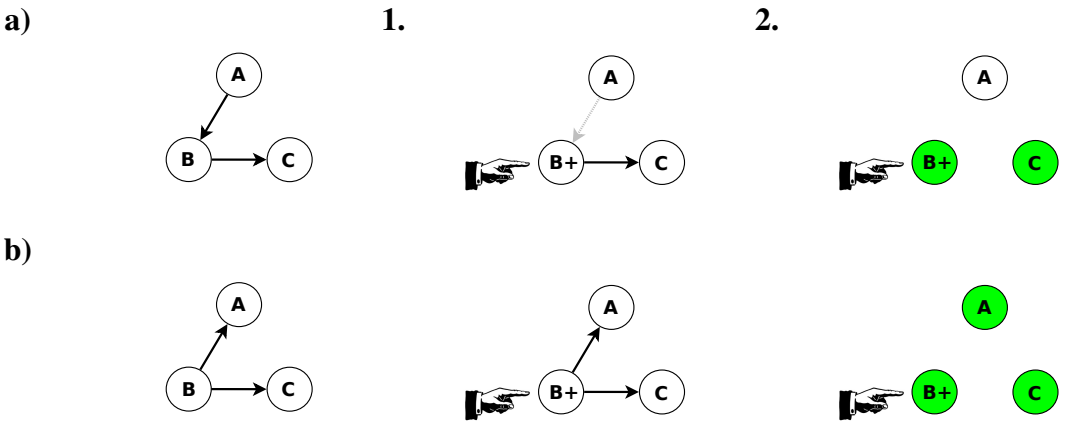


Figure 1.2: Two possible structures a) $A \rightarrow B \rightarrow C$ and b) $A \leftarrow B \rightarrow C$. 1. Intervening on B severs any incoming causal links (dotted in grey). 2. Assuming the goal of distinguishing between these two structures, intervening on B is informative. If we clamp B *on* (+ symbol) and then observe that C also turns on (highlighted in green) but A does not, then we have evidence for structure a) $A \rightarrow B \rightarrow C$. If A also comes on then this is evidence for structure b) $A \leftarrow B \rightarrow C$.

1.1.2 Quantifying interventions

Given a prior probability distribution over structures, interventions generally provide evidence that changes this distribution. Exactly what evidence is provided depends on the intervention's outcome. By summing over the possible outcomes of an intervention, weighted by their probabilities, we can compute their expected values. Assuming known likelihoods, the marginal probability of each outcome $o_i \in O$ given an intervention $q_j \in Q$ and prior $p(G)$ is given by:

$$p(o_i|q_j) = \sum_{k \in G} p(o_i|q_j, g_k)p(g_k) \tag{1.7}$$

Likelihoods for different outcomes $p(o_i|q_j, g_k)$ can be calculated using the known causal powers of the variables, or replaced with an integration over parameter values: $\int_\theta p(o_i|q_j, g, \theta)p(\theta|g)d\theta$ in the general case. The posterior probability of each graph given an intervention-outcome pair is then:

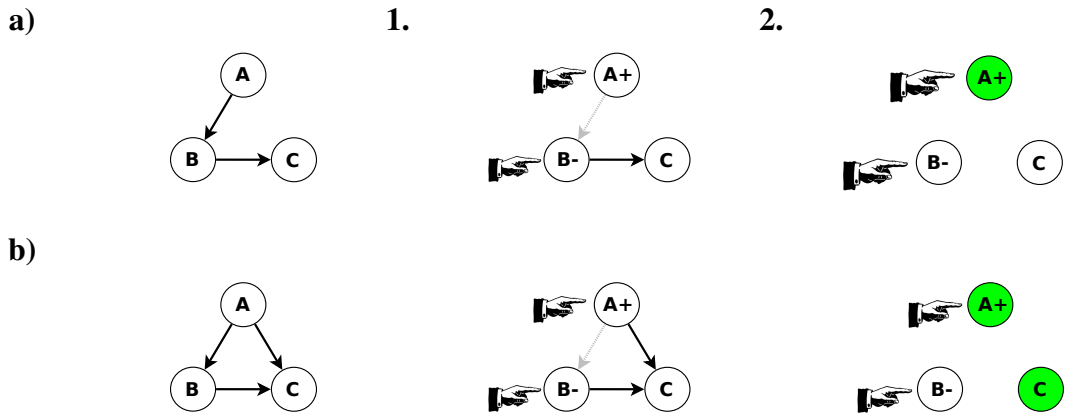


Figure 1.3: To distinguish a) $A \rightarrow B \rightarrow C$ from b) $A \rightarrow B \rightarrow C, A \rightarrow C$ you can manipulate A while holding B constant. 1. If variables are binary, this is achieved by clamping A *on* (+ symbol) and clamping B *off* (- symbol). 2. Then, if C still activates this is evidence for the second, fully-connected structure.

$$p(g_k|q_j, o_i) = \frac{p(o_i|q_j, g_k)p(g_k)}{\sum_{l \in G} p(o_i|q_j, g_l)p(g_l)} \tag{1.8}$$

These can be used to compute any desired summary value $V(G)$ of the resultant distribution. The *expected value* of an intervention is then the sum of these values weighted by the probabilities of the different outcomes:

$$\langle V(G) \rangle_{q_j} = \sum_{i \in O} V(G|o_i, q_j)p(o_i|q_j) \tag{1.9}$$

Intuitively, a good intervention is one which is expected to have a large effect on this probability distribution. Tong and Koller (2001) and Murphy (2001) use expected *information gain* Shannon (1951, Box 1), to quantify the expected values, $\langle V(G) \rangle_Q$'s, of interventions. Steyvers et al. (2003) also used information gain to evaluate interventions in their active causal learning task. There is some evidence that people are able to pick queries that maximise information when active learning about categories (Gureckis and Markant, 2009; Meier and Blair, 2012).

However, information is not the only potentially relevant property of a probability distribution. Alternatively, Kruschke (2001) argues that error reduction is a key driving force of learning, and often guides attention. Reducing the absolute probability of error is equivalent to maximising the probability of the most likely element of the posterior.

In three active concept learning tasks, Nelson et al. (2010) find that participants are better described as acting to maximise expected *highest posterior probability* rather than expected *information*. In another paper (2005), he compares the "usefulness" of different queries according to *information gain* and *probability gain* along with several other measures. He shows that the queries these measures endorse sometimes diverge, dependent on the environment, hypothesis space and prior. Therefore it is very much an open and testable question which measure is a better description of how people select interventions.

Additionally, we have argued that people learn causal representation because it is useful to them; facilitating and simplifying inference. Therefore, the expected *utility* of one's posterior, given some appropriate loss function, can be thought of as the quantity a rational learner should be trying to maximise. The appropriate loss function might often take some idiosyncratic situation specific values , i.e. one might attach much greater significance to being right about the causal structure linking some subset of the variables than one does to the rest of the network. However, a general default loss function is one which treats correct link identifications as having positive utility and mistaken identifications as being worth zero. This formalises the idea that causal structure learning is not an all-or-nothing thing as it would seem to be according to probability gain. Correctly identifying part of a causal structure is often more useful than identifying none of it. Two recent studies have suggested that people are insensitive to outcome utilities when active learning about categories (Gureckis and Markant, 2009; Meder and Nelson, 2012). By explicitly incentivising participants to identify causal links correctly, we can test whether this is the case in causal learning too. Explicit incentives also allow quantification of the performance of information at probability gain methods in terms of maximising task utility.

1.2 Quantifying the value of posterior distributions

1.2.1 Information

A measure of the uncertainty implied by a probability distribution, measured in bits, approaches zero as the probability distribution becomes more peaked.

$$H(G) = -\sum_{k \in G} p(g_k)\ log_2\ p(g_k) \tag{1.10}$$

1.2.2 Highest posterior probability

The probability of the current most likely hypothesis. Goes to 1 as probability distribution becomes more peaked.

$$\Phi(G) = \max_{k \in G} p(g_k) \tag{1.11}$$

1.2.3 Expected utility

Your expected score given you endorse the best possible hypothesis. For each hypothesis, expected score is the weighted sum of the utility of choosing that hypothesis given all possible true hypotheses, j's, weighted by their probabilities.

$$U(G) = max_{i \in G} \sum_{j \in G} L(g_i, g_j) p(g_j) \tag{1.12}$$

An optimal intervention $q \in Q$ with respect to value function $V(.) \in [H(.), \Phi(.), U(.)]$ is thus:

$$argmax_Q \langle V(G|Q) \rangle_Q \tag{1.13}$$

A final issue is that, when learning continues over multiple instances, greedily maximising some expectancy of the posterior (be it information, highest posterior probability, or expected utility) is not guaranteed to optimally maximise that same quantity in the long run. To be truly optimal, a learner should treat the learning instance as a Markov decision process (Puterman, 2009), and look many steps ahead always selecting the intervention which leads to the greatest expected final or total utility (assuming they will maximise on all future interventions). However, computing expectancies over multiple hypotheses when each intervention has many possible outcomes is computationally intractable for all but the smallest number of variables and most constrained hypothesis spaces, dooming any search for strict optimality in the general case (Simon, 1982).

In sum, it is not clear which, if any, of these measures of intervention value can be said to underpin a rational model of intervention selection. Whether maximising information, probability or utility provide better long-run learning efficiency, and whether human active learning is better described by the attempt to maximise one or other of these quantities, are open questions which we explore in the current study.

1.3 Extensions to existing work

As mentioned, the current study and analyses represent significant extensions to the existing literature along several dimensions. We introduce some of these extensions here:

1.3.1 Complex "controlling" interventions

As shown in Figure 3, simple interventions (fixing a single variable) are not sufficient to learn all causal structures. Controlling, e.g. clamping variables off, is vital to identifying certain structures. Moreover, controlling variables is widely seen as a central tenant of science and a key aspect of scientific thinking (Cartwright, 1989; Kuhn and Dean, 2005). The ability to select appropriate controlled tests when necessary, is therefore an important part of successful active learning. To our knowledge, no previously published study has tested whether people are able to select appropriate complex interventions when these are necessary, relying on reduced hypothesis spaces to allow participants to identify the correct structure without controlling variables. Here we allow all 27 interventions (all permutations of variables clamped on and off) instead of just 2-3 single variable interventions. Notably, we also allow interventions in which nothing is clamped (i.e. observations) to test the idea from Steyvers et al. (2003) that people may often use observations to generate hypotheses which they then test with interventions.

1.3.2 Full hypothesis space

Previous studies have always looked at constrained hypothesis spaces, in which to some extent causes and effects come 'prepackaged'. E.g. there are 5, 5, 5-16 and 18 possible models in Lagnado and Sloman (2004), Sobel (2003), Lagnado and Sloman (2006) and Steyvers et al. (2003) respectively. In the general case, there are 25 possible causal structures for three variables, and two cycles[2]. We look at learning over the whole of this space.

[2]We do not want rule cyclic graphs out as psychological causal models just because they do not result in proper Bayesian networks

1.3.3 Multi-shot learning

As mentioned, the only study looking formally at intervention choice (Steyvers et al., 2003), models choice of only a single-shot, one-variable intervention. However many structures take more than one intervention to distinguish. Also, assuming that there is normally some inherent uncertainty in the world, real world causal learners must be able to accumulate evidence over multiple interventions. They should not rule structures completely out or completely in and should be robust to occasional pieces of confounding evidence. Due to their focus one-shot, semi-deterministic interventions, Steyvers et al's study did not explore the rich and probabilistic nature of real world causal learning. To investigate the patterns of intervention selection and belief updating, we allow participants multiple interventions on fully probabilistic structures, and take online structure judgements in order to observe their shifting beliefs.

1.3.4 Minimisation of priors

Previous studies have largely used cover stories to frame the active learning problems. For example, Lagnado and Sloman's (2004) causal variables were levels of chemicals to be combined and Stevers et al's (2003) causal variables were faulty mind-reading aliens. The use of cover stories in causal learning tasks introduces a potential confound. People may base their inferences partly on the prior knowledge that they bring to the problem through its similarity to familiar causal situations or domains (Goodman et al., 2011; Griffiths and Tenenbaum, 2009; Tenenbaum et al., 2006). For example, propensity for participants to select common effect structures might have been due to their interpretation of the (chemical mixing) cover story in Lagnado and Sloman (2004). It is difficult to separate inferential errors from prior beliefs. Our study avoids this as far as possible by keeping the task abstract.

1.3.5 Incentivised learning

A final point is that previous studies did not explicitly incentivise participants. Thus, task demands may sometimes have been ambiguous (should the participant select any causal structure sufficient to explain the data, or must they select the only possible, or most likely, structure?). Without incentivisation there is ambiguity in formal analysis of intervention choices because we cannot assume that participants' goal was to maximise the quantity (i.e. *information* in Steyvers et al) that was used to drive the

ideal learner analysis. Here we reward participants for each causal link they correctly identify. Participants, and ideal learning models therefore have a clear incentive. This approach allows us to model intervention selection and causal structure endorsement as a joint, utility maximising, process, and to compare different methods of intervention selection in a principled way (see Modelling intervention choices).

Chapter 2

The Experiment

2.1 Overview

To test peoples' ability to learn causal structure through intervention, we designed an interactive computer-based active learning task in Flash (see bit.ly/1asZIJE for demo). In the task participants had to use interventions to find and mark the causal links in several probabilistic causal structures. For each structure, participants were faced with three filled grey circles, set against a white background. They were trained that these were nodes, and that they made up a system of binary variables. 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 grey. For each structure participants would perform multiple independent tests before endorsing a causal structure and moving on to the next problem.

Each test had three main stages (Figure 2.1):

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 '+' symbol), clicking again to clamped it to inactive (denoted by a '−' 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, they 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[1]. 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 their 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, anti-clockwise link*, back to *no link*. The initial direction of each link (clockwise or anti-clockwise was randomised). Participants were not forced to mark or update links until after the final test but simply given the option to mark as they went along if desired, as a memory aid.

Pretests insured that participants understood the purpose of the three stages of the task, how to clamp and unclamp nodes, and how to mark causal links.

Participants performed 12 tests on each problem. After their last test, they were prompted to finalise their choice for the causal structure, i.e. they must choose *no link, clockwise link* or *anti-clockwise 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 (Figure 2.1). There were three node-pairs per problem (A-B, A-C and B-C) and three options (*no-link, clockwise*

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

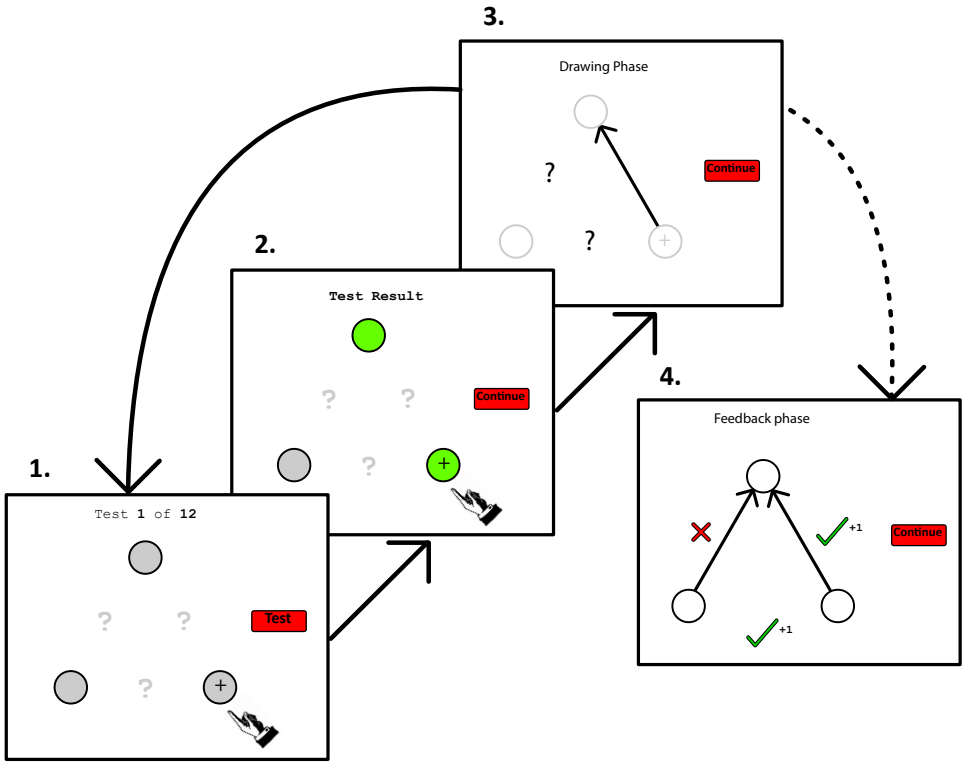


Figure 2.1: The procedure for a problem. 1. Choosing an intervention, 2. Observing the result. 3. Updating causal links. And, after 12 trials, 4. Getting feedback and a score for the chosen graph.

link, anticlockwise link) per node-pair. This means that chance level performance was 1 link correct per problem, or ≈5 points over the five problems while an ideal learner could get 15 or nearly 15 points. At the end of the task participants received $1 plus 20c per correctly identified link leading to a maximum payment of $4.

2.2 Method

2.2.1 Participants

79 participants were recruited from Mechanical Turk. They were paid between $1 and $4 (M=$2.80) depending on their performance.

2.2.2 Design and Procedure

Participants completed one practice problem and five test problems (Figure 2.2). The practice problem was randomly chosen from the five test problems, then the test problems were presented once each, in a randomised order. The location of nodes *A*, *B* and

C were randomised and the nodes were not labelled. Participants performed 12 tests per problem as described above before endorsing a structure and receiving a score. The running score, test number and problem number were displayed across the top of the screen during testing.

Before starting the task, participants were required to correctly answer three out of four questions testing their understanding of the instructions. Participants who did not manage this were sent back to the beginning of the instructions and given one more chance to identify the answers to these questions (52% of those who completed the task). 45 people who got the test questions wrong twice were automatically excluded and did not complete the task.

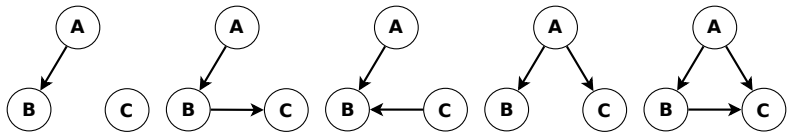


Figure 2.2: Causal test models. From left to right: 1. Single link $A \rightarrow B$, 2. Chain $A \rightarrow B$, $B \rightarrow C$, 3. Common Effect $A \rightarrow B$, $C \rightarrow B$, 4. Common Cause $A \rightarrow B$, $A \rightarrow C$, 5. Fully-connected (chain + direct link) $A \rightarrow B$, $A \rightarrow C$, $B \rightarrow C$.

2.3 Results

2.3.1 Performance

Overall, performances in the task were well above chance. Participants identified an average of 8.97/15 (SD = 4.09) causal links, and got 33.9% of the models completely right. This is well above the chance level of 5/15 links (and 3.7% models correct), $t(78) = 8.60, p < .0001$. The task took around 29.4 minutes (SD = 16.2) to complete and time taken did not have a significant influence on score although there was a trend toward higher scores for faster participants (p=0.07) . The distribution of performances appears somewhat bimodal with one mode at chance and the other near ceiling (see Figure 2.3), suggesting that some participants were not able to solve the task while others did very well. There was no effect of problem order on performance, nor was there an effect of the practice round on when participants faced the same problem again.

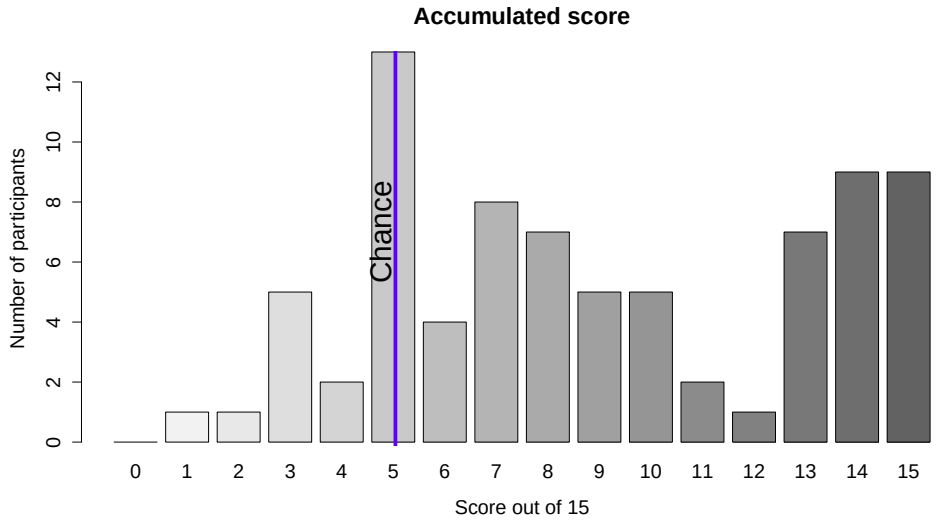


Figure 2.3: Histogram of scores, (5 is chance, 15 is ceiling).

Neither accuracy on test questions, nor whether a participants read the instructions once or twice predicted their performance. Participants found the problems roughly equally hard, with slightly lower scores for chains, common cause and fully connected structures, however there was no main effect of problem on score. They had registered at least one link on 91% of trials and all three links on 76% of these, meaning we have a reasonable record of how their beliefs shifted throughout learning. Participants did not over-connect or under-connect 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%. However, participants did have a tendency to mistakenly endorse the fully-connected structure in place of the chain (Table 1).

Table 2.1: The five most frequent endorsement errors

True structure	Mistaken for	Frequency
$A \to B, B \to C$	$A \to B, A \to C, B \to C$	18/79
$A \to B, B \to C$	$A \to B, A \to C$	7/79
$A \to B, A \to C, B \to C$	$A \to B, B \to C$	7/79
$A \to B, A \to C$	$A \to B$	6/79
$A \to B, A \to C$	No links	6/79

2.3.2 Intervention choice

Singular interventions $A+$, $B+$ and $C+$ were by far the most frequently selected, accounting for 73% of all interventions. This is sensible as these were most often the most informative interventions according to information, probability and utility gain, especially when the learner's prior was relatively unpeaked (see Table 2). On later trials, double interventions (with one variable clamped on, and another clamped off, e.g. $[A+, B-]$) were sometimes more useful. This was usually the case when there were 2-3 equally favoured structures with similar posterior probabilities. The common cause, common effect and singly linked structures can be distinguished using only single interventions. Participants selected the 6 informative double-interventions 6.1% of the time, and were more likely to do so when faced with chain or fully-connected structures (10.0%), and did so more on later trials ($t(272) = 5.90, p < .0001$). Propensity to use these interventions was strongly predictive of performance. For each additional double intervention performed, participants scored .22 extra points over the task ($F_{1,77} = 8.053, p = 0.0058$), this effect was even stronger looking specifically whether participants used these when learning the chain or fully connected structures ($F_{1,77} = 20.41, p < .001$). There were 7 relatively uninformative, dominated interventions (leaving everything unclamped, clamping a single variable off or clamping two variables on) which provided very little information and were never the most informative choices to make. These were nevertheless selected 13.4% of the time, but much more often by the poorer performers ($F_{1,77} = 26.84, p < .001$, -.18 points per additional dominated intervention). Finally there were 11 completely useless interventions, which provided no information whatsoever (all triple interventions or clamping two variables off) but which were still selected 7.6% of the time. Again, these were much more likely to be selected by poorer performers ($F_{1,77} = 19.16, p < .001$, -.22 points per additional useless intervention), with participants who selected one of these interventions at least once averaging only 6.39 points over the task, which is not significantly greater than chance. Participants almost never chose to leave everything unclamped and just observe (.06%).

The interventions selected by participants were generally informative. Participants got much more information than they would by picking interventions at random (see Figure 2.4) averaging 2.77 times the information, and raising their expected utility 2.31

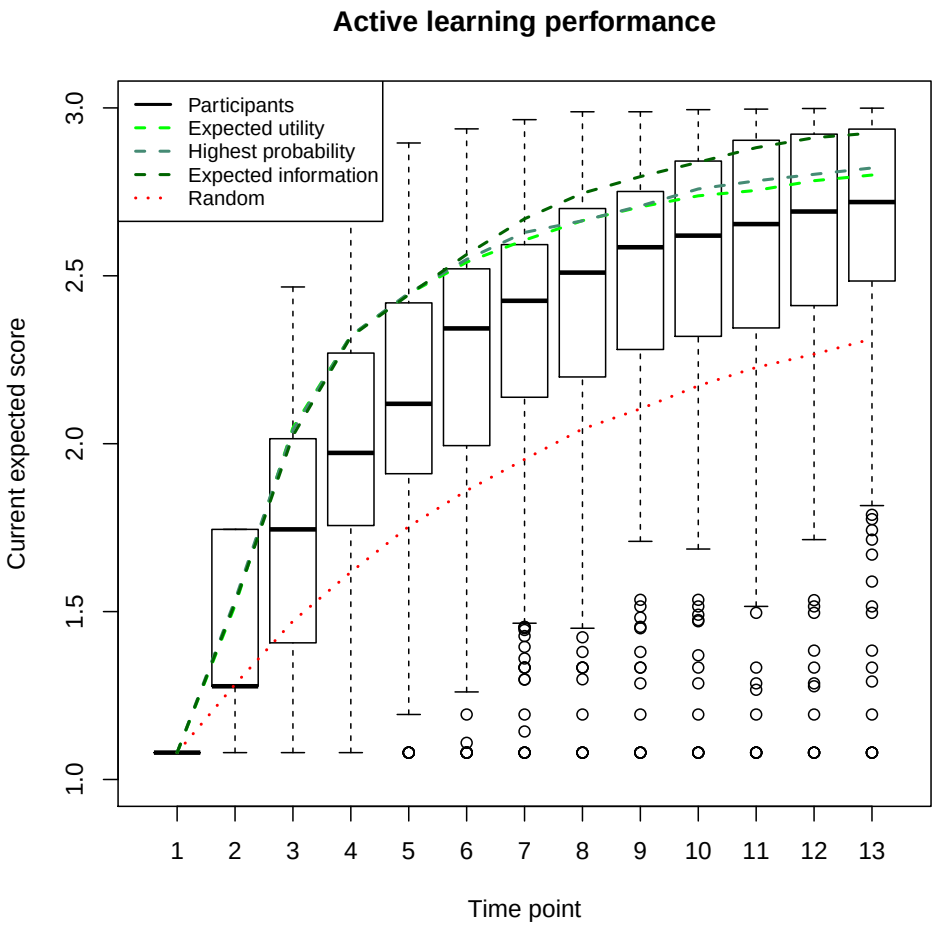


Figure 2.4: Ideal *expected score* at each time point given each participant's interventions from t=1:t. For comparison lines are included for average expected score given random interventions (red) and given ideal utility, probability and information maximising interventions (greens).

times faster, than they would by selecting interventions at chance. On the other hand, they were not optimal, averaging .50 of the information gain, and .33 of the utility gain of those chosen by an ideal learner. Participants finished problems having gathered enough information that they could optimally expect to score 2.63 (SD=0.59) points (13.15 over the whole task), this was significantly higher than selecting interventions at random which would permit only 2.31 points ($t(394) = 16.21, p < 0.0001$). However this was also considerably lower than what could be achieved by intervening to maximise information, probability or utility. These approaches averaged expectancies

of 2.92, 2.82 and 2.80 points per problem respectively[2]. For comparison, merely observing the system without clamping any variables would have provided very little information, allowing 1.87 points per problem on average. The quality of participants' interventions were also strongly positively correlated with their ultimate performance. This is true for all measures of intervention quality tested here: Information gain: $F_{1,77} = 33.65, R^2 = 0.30$, probability gain: $F_{1,77} = 42.36, R^2 = 0.35$, and utility gain: $F_{1,77} = 27.31, R^2 = 0.26$ (all p's $< .001$, see Figure 2.5).

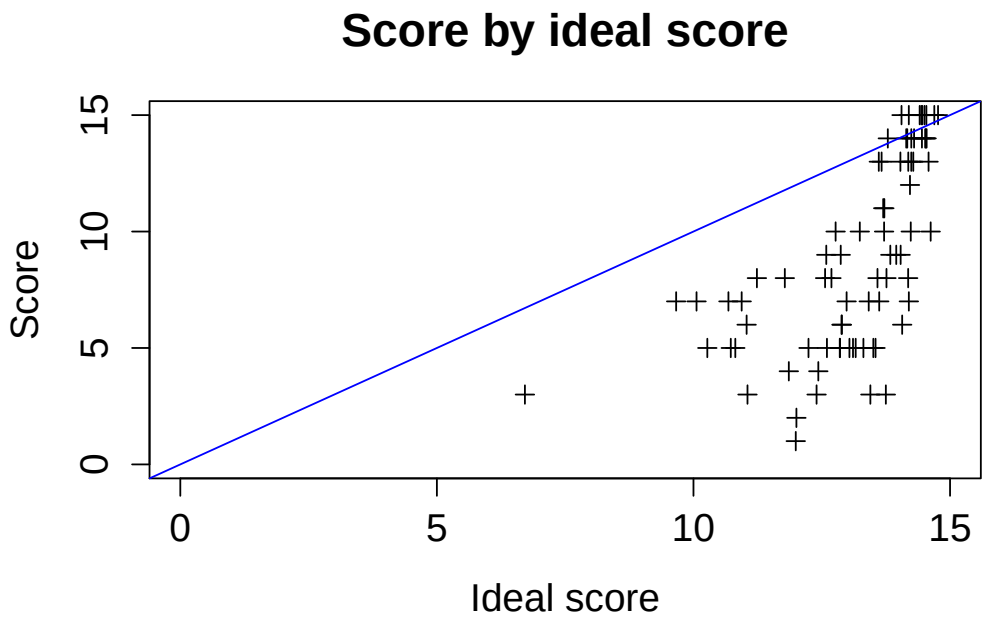


Figure 2.5: The score achieved by participants against their ideal score, which assumes perfect integration of the information and maximisation of utility. An optimal performance given a set of interventions would lie on the blue line. The more useful the set of interventions selected by the participant, the further to the right their point lies.

2.3.3 Optimal performance?

Within 12 trials it was possible for an efficient learner to approach a perfect performance, averaging at least 14 points depending on the choice of intervention values being used in learning (14.6 information, 14.2 probability, 14.1 utility in the simulations mentioned above). 23% of participants scored 14/15 or 15/15 suggesting a significant

[2]Figures based on simulating the task 1000 times per intervention choice rule, see bit.ly/13dmcVw for code.

minority of participants were very close to these optima. The three value functions made similar but occasionally diverging choices. The vectors of values for expected information, probability and utility were fairly well correlated with bivariate correlations of information-probability = .78 , information-utility = .79 and probability-utility = .84, but often did not share maxima: 61%, 64% and 82% of the time respectively. Part of this discrepancy stems from later trials, when expected utility and probability failed to distinguish between interventions, with no expected increase for any intervention for highly peaked probability distributions. This inability to distinguish has been noted as a problem for probability gain before (Jonathan Nelson, personal correspondence). Looking two steps ahead, all three measures converge somewhat although there are still discrepancies. Correlations increase to .84, .92 and .98 and the measures have the same maxima on 90%, 88% and 83% of participant trials. Looking two steps ahead, the ideal learner models average almost identical performance (14.6, 14.4, and 14.6 points respectively), suggesting there is little more to gain by looking further ahead in this environment.

2.4 Interim discussion

Overall performance levels are in line with, or slightly higher than, those found in past studies. After performing fifty interventions, 35% of participants identified the right structure out of a five alternative forced choice in Lagnado and Sloman (2004). This was greater than the performance in their observation condition, but was only marginally greater than chance. In Steyvers et al's (2003) study, 33% (lab) / 34% (web) of structures were correctly identified after a single intervention from an 18 alternative forced choice. In our study, participants also correctly identified 34% of structures correctly despite the larger hypothesis space of 27 (all possible 3-variable belief networks plus the two cycles $A \rightarrow B \rightarrow C \rightarrow A$ and $A \rightarrow C \rightarrow B \rightarrow A$). With the exception of Lagnado and Sloman's (2006), experiment 3, none of these past studies were fully probabilistic. In the first two experiments of Lagnado and Sloman (2004) and the relevant experiment in Lagnado and Sloman (2006) effects never occur without their causes but causes can occur without their effect. In Steyvers et al., the intervention manipulation set a variable to a state that none of the variables would normally take. If another variable then took on this state, it provided unambiguous evidence for

the presence of a link. This means that in all of these studies, sensible interventions made it possible to be perfectly sure of the true structure. This was not the case in our study. No intervention could perfectly rule a structure completely in or out. Lagnado and Sloman (2006) experiment 3 was fully probabilistic but, with a 16 option forced choice, performance was very low, ranging from 17% correct to 4% correct depending on the structure being learned. Thus, our study was both harder and more general than past studies, while performances were comparable or better. 23% of participants performed near-optimally (14/15 or 15/15 links identified), which is comparable to the ideal learner simulations, suggesting motivated causal learners can be capable of accurate and sophisticated inference.

It was slightly surprising that there were no significant differences in performances between structures. Lagnado and Sloman's (2004) experiments suggested people may have a preference for selecting a common-effect structure rather than a chain when activations take place simultaneously, as they did in our task. This might have led to better performance on our common effect structure but was not the case here, suggesting that the bias identified by Lagnado and Sloman was due to their cover story. However, we did find a tendency for chains to be mistaken for fully-connected structures (18 out of 79 participants, compared to 20 out of 79 who correctly identified the chains). This is in line with Ferbach and Sloman's (2009) study which suggests that people may often infer individual causal links based on local computations. Fernbach and Sloman's heuristic model adds direct links from intervened on variables to activated variables. By doing this it avoids the complexity of updating a large hypothesis space, but leads to systematic mis-attribution of chains as fully connected structures. This is because, upon intervening with A^+, both B and C generally activate, leading to the attribution of direct links $A \rightarrow B$ and $A \rightarrow C$. When B^+ is tested, C tends to activate leading to addition of $B \rightarrow C$. To distinguish these models, a double-intervention $[A^+, B^-]$ is required (Figure 1.3). In our study we find that the best performers do manage to select and learn from double "controlling" interventions, indicating that a local computation heuristic does not tell the whole story.

Steyvers et al. (2003) proposed that people might use observations to generate hypotheses which they then test through intervention. However, there was no sign, in our experiment, that people choose to merely observe (test with no nodes clamped) if

they can perform an intervention for no more cost.

The apparant bimodality in performances is in line with Steyvers et al who found that they could separate participants' causal inferences (in their observation only experiment) by model-based clustering. Their analysis split participants into one group (42%) whose inferences were at chance, and one to two groups who were sensitive to the evidence but differed in their ability to remember past trials (24% remembered evidence from past trials and 34% based inferences purely on the final trial). By extending over multiple problems and multiple interventions, our study provides more compelling evidence that while some people can achieve very high and robust active causal learning performances – the probability of getting 14/15 or 15/15 in this task by chance are tiny (around 2×10^{-7}) and yet 23% of participants achieved this – others remain clearly at chance.

Intervention choice was correlated with performance, with selection of more useful interventions indicating better overall performances. However, this correlation is partly a necessary consequence of intervention choice, i.e. an ideal passive learner could not expect to exceed the ideal score available given a set of intervention choices. However, in this task there is clearly more variance in score than in ideal score(Figure 2.5) suggesting that good quality interventions are a necessary but not sufficient aspect of successful active causal learning. We presume that a large part of what intervention choice can tell us about how people achieve causal learning will come from what it reveals about the inferential processes people are engaged with when solving the task.

Previous studies have approached this question by including a yoked condition, where another set of participants are shown, or forced to perform, the interventions selected by the participants in an intervention condition, thus removing any information asymmetry. If intervention quality fully drove performance we might think that there should be no difference between these groups. These studies have had mixed results, sometimes finding a difference between conditions (e.g. Sobel and Kushnir, 2006) and sometimes not (e.g. Lagnado and Sloman, 2004). However, there are many reasons why yokees might perform differently to their yokers. These include mundane differences in attention and motivation, but also more complex biases due to differences in the hypotheses being considered by the learners as the information is presented to them (Markant and Gureckis, 2013). Due to the large hypothesis spaces and the complex

way in which new evidence affects the posterior, comparing to a yoked condition is unlikely to support strong conclusions about peoples' causal learning processes. For this reason we have opted for a different approach, explicitly fitting intervention-attribution models to the actions and model endorsements of individual participants. This approach allows us to ask more sophisticated questions of the data. Specifically, we can ask what latent variables drive performance differences between individuals; what criteria learners use to select interventions; or whether there is evidence of systematic heuristic use.

Chapter 3

Modelling participants' active learning

In the following section we will describe a number of models which we fit to individuals data using maximum likelihood[1]. The trial-wise data from this task is multivariate. For each trial the participant chooses an intervention from the 27 legal interventions but also gets to update their marked causal links from the 27 possible combinations of causal links. Each model assigns a probability to each intervention and to each combination of marked (and unspecified) links on each of the 12 trials for each of the five problems. Free parameters are fitted to individuals. We do this because we assume that properties like memory and learning strategy are fairly stable for each individual but likely to differ between participants in ways which may help us understand what drives the large differences between their performances. Models are compared using their Bayesian information criterion (*BIC*, Schwarz, 1978), which provides a conservative approximation of their posterior probabilities.

3.1 What drives our intervention choices?

Earlier in the paper (Box 1) we introduced several measures of the value of a probability distribution which differed in their usefulness in driving active categorisation (Nelson, 2005; Nelson et al., 2010), but which all turned out to be reasonable methods for choosing useful interventions in the current task environment (see *Optimal Performance?*). The measures make different predictions about the value of the possible interventions dependent on the current prior. This means that we can look to see whether participants are better described as intervening to maximise one or the other. We will

[1] We used the Nelder-mead algorithm as implemented in R's "optim" function. Optimisation was validated by repetition with different starting parameter values.

call a model which assumes people try to minimise their uncertainty (without worrying about their probability of being right, or how much they will get paid) a *scholar* model. We will call a model which assumes people are just concerned with maximising their probability of being completely right (disregarding all other possible outcomes, or their payouts) a *gambler* model. Finally a model which assumes participants are pure pragmatists, trying to learn only in order to maximally increase their expected score, a *utilitarian* model.

We begin the model fit by temporarily assuming that participants are able to optimally update their subjective probability distribution over possible causal structures. Thus, we take each probability distribution $P(G)^t$ to be the true posterior, starting from a flat prior at $t = 0$ and updating with the outcome of each test up to time point t.

3.1.1 The scholar

This model assumes that participants choose actions to maximise their expected *information*, or equivalently minimise their expected *uncertainty*, about the true structure, denoted by $\langle H(G) \rangle$ (see Box 1).

3.1.2 The gambler

This model assumes participants choose actions which maximise the expected posterior probability $\langle \Psi(G) \rangle$, of the most likely structure.

3.1.3 The utilitarian

This model assumes that participants choose actions which maximimise their expected payout, $\langle U(G) \rangle$ with each intervention. Expected payout was calculated by creating a matrix of payouts for the 27 possible endorsements (including the two loops), given the 25 directed acyclic graphs, and multiplying the appropriate line with the posterior distribution. The cells of the matrix were populated according to the payouts received if one endorsed graph i, given graph j was the true graph. For example, the cell for 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 while the other two are wrong.

While we are not yet capturing the idea that people might not be perfect Bayesian updaters, we can capture the idea that there is likely to be some error in their maximisa-

tions over intervention values. We assume that each learner soft-maximises over intervention values V to a degree controlled by a parameter α. Let $v_1 ... v_n$ denote the values $\langle H(G|Int_{1..n})\rangle$, $\langle \Psi(G|Int_{1...n})\rangle$ or $\langle U(G|Int_{1...n})\rangle$ depending on the model. The probability that the learner selects intervention i at time t is given by:

$$P(\text{Int}_t = i) = \frac{e^{\alpha\langle v_i\rangle_t}}{\sum_{k=1}^{n} e^{\alpha\langle v_k\rangle_t}} \tag{3.1}$$

This means that, if $\alpha = 1$ the probability that the learner picks an intervention is proportional to its value relative to the other interventions (akin to probability matching). As $\alpha \to 0$, the value distribution over interventions approaches uniformity, since the probability of picking intervention i approaches $e^0/ne^0 = 1/n$. However, for α's greater than 1, as $\alpha \to \infty$ the probability that the learner picks the most valuable intervention approaches 1, and the probability of picking the other interventions drop away toward 0.

The probability that the learner marks a particular combination of causal links j at time t is similarly considered to be a soft-maximisation over the true posterior distribution fit to individuals with a second free parameter β:

$$p(\text{Stated-belief}_t = j) = \frac{e^{\beta p(g=j)_t}}{\sum_{k=1}^{m} e^{\beta p(g=k)_t}} \tag{3.2}$$

The marking of links on each trial was optional, and initially all links were unspecified. Therefore, on many trials participants had not fully stated a single causal model but rather a set of models q consistent with the stated edges j. To capture this, the models marginalise over all structures consistent with the marked links on each trial:

$$p(\text{Stated-belief}_t = j) = \frac{\sum_{q=1}^{j} e^{\beta p(g=q)_t}}{\sum_{k=1}^{m} e^{\beta p(g=k)_t}} \tag{3.3}$$

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 which consistent with this link. If a participant has not marked any links then this trivially has a probability of 1[2]. On the 4.3% of trials in which participants marked a cyclic

[2]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.

structure ([A→B, B→C, C→A], or [A→C, C→B, B→A]) their belief state was treated as unspecified.

3.1.4 Far-sighted scholars, gamblers and utilitarians

As mentioned in *Intervention choices*, the values of different actions are similar but not identical if the learner looks further into the future. Computing expected values looking more than two steps ahead becomes impossibly slow even in the three-variable case, but we were able to compute the two-step-ahead expectancies for the three measures[3]. This allows us to check if there is evidence that people are able to look more than one step ahead when choosing interventions. Accordingly three additional models were fit in which the vectors of intervention values for looking one step ahead were replaced with those looking two steps ahead.

Based on this first round of model fits the scholar model appears to provide the best overall fit. Individually, the large majority of learners are better described as scholars than as gamblers or utilitarians (see Table 3.1 and bit.ly/13dmcVw for full results). The β parameters are identical across models within participants because all of these models assume that participants maintain a Bayesian posterior over hypotheses throughout the task, meaning this parameter was unaffected by the interventions chosen. However, the α parameter for the Scholar model, controlling the degree of maximisation over intervention values, varies considerably between participants and strongly predicts performance with an increase of .82 points for every 1 increment of α ($F_{1,77} = 104.31, p < .0001, R^2 = .56$).

The five participants better fit by the *gambler*, *utilitarian* and *farsighted gambler* models had an average score of 6 which is not significantly above chance. Including the 6 purportedly *farsighted scholars* this is still only 6.27.

3.2 Bounded models

3.2.1 Forgetful learners

The analysis in Steyvers et al. (2003) suggested that ability to remember evidence from past trials is limited and that differences in forgetting rate explained performance. Many

[3]These expectancies are computed recursively, taking the max over the second set of interventions and passing these values back to the first set of interventions. See http://bit.ly/13dmcVw for code.

Table 3.1: BIC's and median parameter values for initial models. Number (♯) of participants best fit by each model. Best fitting overall model according to BIC in bold.

Model	Steps-ahead	BIC	α	β	♯ participants
Scholar	**1**	**45047**	**5.28**	**3.54**	**68**
Gambler	1	47917	9.09	3.54	2
Utilitarian	1	47464	5.36	3.54	1
Scholar	2	45525	3.48	3.54	6
Gambler	2	46877	11.7	3.54	2
Utilitarian	2	46778	7.69	3.54	0

of their participants chose models that suggested they remembered only the result of their final intervention (having forgotten the evidence from their previous observations) while others rembered a little more. This is in line with what we know about the very limited capacity of working memory (Cowan, 2001; Miller, 1956) and its close relationship with learning (Baddeley, 1992). A simple way to work a memory constraint into our models is to assume that real learners forget a certain amount of their posterior after each trial. This is achieved by including a forgetting rate parameter which is applied to the subjective posterior distribution over structures for a given subject after each trial. The posterior probability distribution is "flattened" at each time point to an extent controlled by a parameter γ. This is achieved by mixing in a uniform distribution, to create a new distribution $P^*(G)$:

$$P^*(G)^t = (1-\gamma)P(G)^t + \gamma\frac{1}{m} \tag{3.4}$$

where m is the number of causal structures. This affects the values of the interventions at the current, and subsequent time points, coupling the three model parameters together. This occurs because each maximisation over interventions and beliefs now takes place over a different, flatter, subjective probability distribution.

We found that the forgetful *scholar model* provided the best overall fit. 75/79 participants were better fit by a *forgetful* model than by any of the initial six models. The majority of these are again better described as scholars (40) but now there are also 20 better described as utilitarians and 16 as gamblers. A potential explanation for this shift is that the forgetful utilitarian and gambler models may be better at fitting random

noise participants. 50.0% of those classed as forgetful utilitarians or gamblers scored at or below the chance level of 5 points while this was true of only 7.5% of those best described as scholars (giving a count of 37 forgetful scholars, 8 forgetful gamblers and 10 forgetful utilitarians, excluding those who performed at or below chance). The participants better described as gamblers or utilitarians had extremely high forgetting rate parameters (99.9% and 98.2% respectively compared to the 71% of those described as scholars) meaning that their fits were to a large extent driven by the distributions of intervention values coming from a flat or nearly flat prior. The three simple interventions (A^+, B^+ and C^+) were selected very frequently by all participants, and the distributions of values for probability and utility gain given a flat prior are more peaked for these choices than for information gain. Therefore model fits in which forgetting rate λ approaches 1 (100%), favour the gambler or utilitarian models while not being particularly theoretically interesting. Drawing strong conclusions participants for whom we have no evidence of their performing differently from chance, based on models with very extreme parameters, may be misleading, so in subsequent analysis we present both the absolute number of participants best fit and the number including only those who scored above 5 points.

Degree of forgetfulness γ (assuming the Scholar model) was strongly predictive of performance ($F_{1,77} = 75.75, p < .001, R^2 = .48$) with each additional 10% remembered per trial leading to an additional .88 points over the task.

3.2.2 Conservative and forgetful learners

Given that even successful participants appeared to forget such a large percentage of the evidence they received in prior trials (median 71% 'forgetting' parameter λ per trial for scholars), it is perhaps surprising that many of them are able to achieve high levels of accuracy over the task. One possible explanation (Harman, 1986) is that people may mitigate their forgetfulness about old evidence, by remembering just what they have previously concluded from it. For example, suppose a participant registers an A→B causal link after their first three interventions. We can take this as a (noisy) indication they are fairly confident 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 their sixth intervention, they might not remember why they had concluded that there is an A→B link three trials

earlier, but they would still be sensible to assume that they 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 one has already registered than those that are inconsistent.

This can be modelled by adding a step which takes place before a learner chooses their next intervention. This step creates a temporary *conservative* probability distribution $P^{**}(G)$ in which causal structures consistent with the currently marked causal links are up-weighted by a factor of η and the whole distribution is renormalised. The conservativeness parameter only does work when participants have registered their beliefs about at least some of the link-spaces, but this is the case on 91% of trials. On 76% of these trials participants had registered a belief for all three link-spaces meaning that the conservativeness parameter up-weights the subjective probability of this one structure while computing the next intervention. This means that even if this learners' posterior is relatively flat due to forgetting, their marked structure still stands out, leading them to behave as if they have selectively remembered information confirmatory about this hypothesis. On the 24% of of trials in which some but not all links remained unspecified, the conservativeness parameter led to the structures consistent with the established links (2-3/25 or 8-9/25 depending whether one or two links remained unspecified) being up-weighted, leading the learner to favour interventions likely to distinguish between these options. Formally, we use a vector δ of length m whose entries are zero at all causal structures inconsistent with the current selection of causal links, and one otherwise:

$$P^{**}(g_i)^t = \frac{\eta^{I[\delta_i=1]} P^*(g_i)^t}{\sum_{j=1}^{m} \eta^{I[\delta_j=1]} P^*(g_j)^t} \tag{3.5}$$

The model then uses this temporary altered distribution to compute the expectancies of different interventions and soft-maximises over these according to degree α as before. The conservative distribution $P^{**}(G)^t$ is then discarded, and $P^*(G)^t$ is updated according to the interventions' outcome to give $P^*(G)^{t+1}$, whereupon the process is repeated.

The fit of the conservative and forgetful scholar, gambler and utilitarian models are given in Table 3. The overall best fitting model was the conservative forgetful scholar model with the lowest overall BIC score of all the models we fit, (pseudo-$R^2 = .44$)[4].

[4]McFadden's pseudo-$R^2 = 1 - \frac{ln(M_{full})}{ln(M_{minimal})}$. Here the minimal model assumes all actions and beliefs are randomly chosen. Values between .2 and .4 are considered a good fit (Dobson, 2010).

Table 3.2: Overall BICs and median parameter estimates for forgetful and conservative models. ♯ participants best described, and ♯ excluding those who did not score above 5 points.

Model	BIC	α	β	λ	η	♯ par (/79)	♯ par$_{>5}$ (/57)
Forgetful Scholar	37840	7.72	23.9	.717		0	0
Forgetful Gambler	40228	40.5	77.4	.946		0	0
Forgetful Utilitarian	39224	12.2	38.1	.894		0	0
Conservative Scholar	40814	6.22	4.76		48.4	0	0
Conservative Gambler	43863	10.4	4.66		18.5	0	0
Conservative Utilitarian	43231	6.83	4.71		32.7	0	0
Conservative Forgetful Scholar	**31145**	**7.87**	**86.1**	**.909**	**2.15**	**31**	**29**
Conservative Forgetful Gambler	31575	61	374	.994	1.35	29	19
Conservative Forgetful Utilitarian	31630	13.4	245	.979	1.59	14	8

74/79 of people were better described as conservative and forgetful rather than either or neither. Again the largest number of individual participants were best described by the scholar compared to the gambler or utilitarian. This trend is clearer if one discounts fits to participants at or below chance.

Looking at the parameters (assuming the scholar model), more forgetful people were also more conservative ($\lambda \propto log(\eta), F_{1,77} = 4.742, p = .0325$) but there was not quite a significant effect of conservativeness on score ($p = .086$). Once again, forgetfulness was strongly correlated with score ($F_{1,77} = 21.4, p < .001, +.25$ points per each 10% less forgetful)

3.3 Heuristic models

We can also consider 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 et al., 1999).

3.3.1 The simple endorser

Fernbach and Sloman (2009) argue that people often identify causal models by treating each link as independent and using single interventions as local cues. Concretely, this theory suggests people simply clamp variables on, one at a time, and tend to add links to any other nodes that activate as a result (tending to remove any links to other nodes which do not activate as a result). We can operationalise this with a model which selects one of the simple interventions with probability θ or else selects anything else with probability $1-\theta$. With a probability σ the belief 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 ϱ) or takes any other state (with probability $1-\varrho$). The probability of an intervention is then:

$$P(\mathrm{Int}^t = i) = \frac{I[Int^t = A^+ \vee B^+ \vee C^+]}{3}\theta + \frac{I[Int^t \neq A^+ \vee B^+ \vee C^+]}{n-3}(1-\theta) \quad (3.6)$$

and the probability of a registered belief state is:

$$P(B_t = i) = I[B_t = B_{t-1}+L]\sigma + I[B_t = g_{t-1}](1-\sigma)\varrho + I[B_t \neq (B_{t-1} \vee B_{t-1}+L)]\frac{(1-\sigma)(1-\varrho)}{m-2} \quad (3.7)$$

It is interesting to note that this behaviour is a limiting case of the expectancy based models. When forgetting rate approaches 100%, all three expectancy based models endorse selecting single variable clamps, and when other variables activate as a result, this leads to a big spike in the probability of any models containing direct links from the clamped variable to the activated variable. This is because singly clamped interventions A^+, B^+ and C^+ are the most informative whenever the prior is relatively flat. However, when both other variables activate, the expectancy based models begin to diverge, predicting that participants should raise their degree of belief in all models which would lead to a double activation (chain, common cause and fully connected) more or less equally, while simple endorsement expects only the common cause structure to be more likely to be registered by participants.

3.3.2 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 structure (Cartwright, 1989; Kuhn and Dean, 2005). In this experiment this takes the form of a double clamp 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 turn clamp one node on and both other nodes activate). This action tests whether the node which remains unclamped is a direct effect of the node which is clamped (Figure 1.3), 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 un-clamped and the other $N - 2$ nodes would be clamped off.

The model is operationalised as selecting A^+, B^+ or C^+ or a disambiguation step (e.g. A^+B^- etc) with probability θ and something else with probability $1 - \theta$. Propensity to select a simple endorsement step rather than a disambiguation step is governed by parameter κ. If a the disambigation step is performed and the unclamped node does not activate, then any link L from the activated node to the unclamped node, is removed with probability σ. In other respects the model is identical to the simple endorser.

Table 3.3: BIC's and median parameter values for heuristic models. ♯ of participants best fit by each model. Best fitting overall model according to BIC in bold.

Model	BIC	θ	κ	σ	ϱ	♯ par
Simple endorser	37334	.850		.588	.233	40
Disambiguator	**36938**	.933	.982	.588	.233	39

Both models do fairly well. Their overall BICs are in the range of the *forgetful* models, with half of the participants described better as simple endorsers and half better as disambiguators.

As expected, those better described as disambiguators have much lower forgetfulness parameters (according to the scholar model; $F_{1,77} = 6.02, p = .016$) with disambiguators forgetting 12% less per trial than simple endorsers. There is no significant difference in performance between the two groups. Looking only at those also

best described as scholars there is a trend for disambiguators to score higher, but this difference is not significant ($F_{1,29} = 2.57, p = .11$).

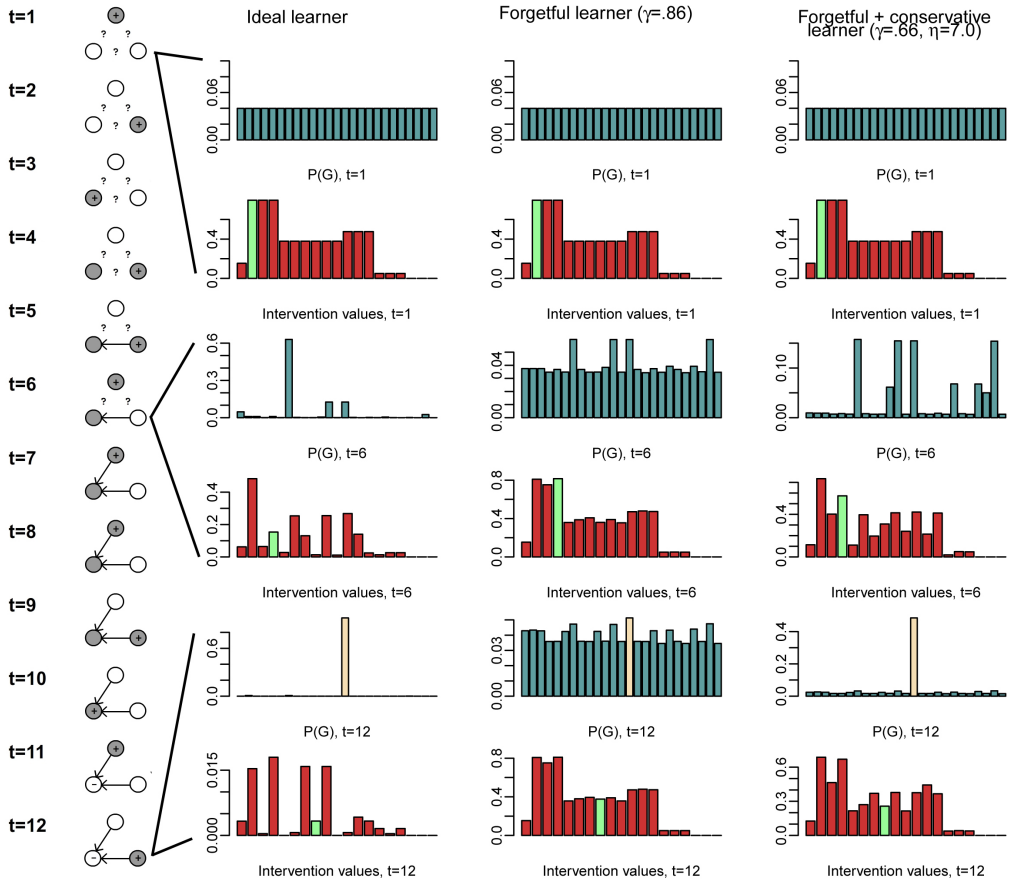


Figure 3.1: Visual comparison of fit for *ideal, forgetful* and *forgetful + conservative scholar* models. LHS: Participant 24, identifying the common effect ($A \rightarrow B, C \rightarrow B$) structure. "+" and "−" symbols indicate participant's interventions, grey nodes indicate the resultant activations, and the arrows replicate those marked by the participant at each time point. RHS: Participant's subjective probability distribution over graphs (blue), and relative values of the different interventions (red) at t=1, t=6 and t=12, according to the three models. Graphs and interventions are ordered loosely by by complexity (i.e. the more links, or more variables fixed, the further to the right). Fully-clamped interventions are excluded to save space since they carry no information. Intervention selected by participant 24 marked in green, and their final structure judgement marked in off-white. Note the comparatively flat posterior distributions at t=6 and t=12 for the *forgetful learner*, and the peaks at the graphs consistent with the $C \rightarrow B$ at t=6 for the *forgetful + conservative learner*, and with the final graph at t=12.

Chapter 4

General Discussion

We have compared computational models of efficient causal learning, driven by three plausible measures of interventions values: expected information (*scholar* models), highest posterior probability (*gambler* models) and expected utility (*utilitarian* models). In all of our comparisons, *scholar* models, driven by information gain, tended to beat those driven by error reduction/probability gain (*gamblers*) or by the objective utilities (*utilitarians*). This was particularly clear looking at the ideal learner models (with no 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 maximising information rather than that of minimising error or maximising utility.

These findings are support what Markant and Gureckis (2012) and Meder and Nelson (2012) found for active classification. People do not, or cannot, worry about outcome utilities when actively learning causal structure. The better fit for the *scholar* than the *gambler* models is in line with what Meier and Blair (2012) found but opposite to what Nelson et al. (2010) find for classification. People seem to be more concerned with reducing their uncertainty than minimising their chance of error.

In a more abstract sense, one can argue that maximising information does provide a good solution to the computational problem of efficiently maximising utility. In our simulations we found that maximising expected utility with each intervention is, paradoxically, not the best way to maximise expected utility over the whole task (see Figure 2.4). One can actually do a little better by focusing on increasing one's information on each step than one can by focusing on one's expected utility or probability of being correct. This has been noted before. For example, information can be shown

to be a much more efficient criterion than probability in a binary feature classification game like Guess Who (Markant et al., 2013). Differences in this task almost evaporate if one looks two steps ahead, but there is no evidence that participants were able to do this, with across-the-board worse fits for models which assumed participants could maximise over the two-step-ahead expectancies instead of one-step-ahead (Table 3.1). Therefore, our modelling suggests that in as far as people evaluate interventions' usefulness explicitly, they are "informavores" (Miller, 1984) intervening to glean as much information as possible, without worrying about outcome factors like probability of error or expected utility.

Venturing one rung down the ladder leading from the computational level toward process level plausibility (Jones and Love, 2011; Marr, 1982), we also explored bounded versions of these models. We included parameters capturing the idea that people might calculate these expectancies as best they could given plausible memory and representational space constraints. We found much better fits by including forgetfulness and a conservatism parameters, but also that the scholar-gambler-utilitarian distinction became somewhat less clear (though overall fits still favoured the scholar model). Including both forgetfulness and conservatism led to much better overall fits than including only one or the other. The two parameters were correlated, suggesting that they complemented one another: the more forgetful a learner is about past evidence, the more conservative they need to be about their beliefs.

It is, of course, possible that allowing participants to draw and update models as they went along will have affected how they approached the task, distracting them, or leading them to place more emphasis than they otherwise would have down on earlier marked links. And, it is true that the models participants drew at each time point are at best noisy markers for their beliefs about the true structure. However these are largely inevitable aspects of tracing beliefs throughout learning. We tried to minimise the extent to which eliciting beliefs distracted participants by making the step optional, and hoping that participants would voluntarily record their beliefs, choosing to take advantage of this stage as an aid to memory. To a large extent it seems this was what participants did, as links were drawn on 91% of trials and neither varied wildly nor remained static from trial to trial. Therefore, this round of models provides a plausible account of how moderate forgetting and a degree of conservatism can lead to good

overall performance and approximation of ideal scholarly behaviour.

Venturing 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 and Sloman, 2009), based on local computation, could capture much of the behaviour of many participants. This may explain why so many participants endorsed the fully connected structure when the true structure was the chain (see *Intervention choice*). However, importantly, some participants also performed the crucial controlling disambiguation steps which cannot be easily captured in a local computation framework. We operationalised this as an alternative step, occasionally performed at random. One way of thinking of the way these two steps work together is that the simple endorsement steps tend to connect up a causal structure, while disambiguation steps potentially prune out unnecessary links (see Figure 1.3). We found that these two simple models performed well, better than the ideal learner models but not as well as their bounded counterparts. Whether someone was classed as a simple endorser or a disambiguator was linked to how forgetful they were according to the bounded models, with less forgetful participants more likely to be classed as disambiguators. This makes sense since, as we noted in *Intervention choice*, disambiguation steps were not the most informative actions in this task when ones prior is relatively flat (as it tends to be for the more forgetful participants). The disambiguator model has the potential to be refined by adding sequential dependence: assuming disambiguation steps are likely to follow ambiguous evidence (i.e. multiple activations) and that learners will generally pick a disambiguation step with the same node clamped on as they had clamped on at t-1.

4.1 Future work

An interesting avenue for future work could be to look at the range of environments in which such heuristic strategies are effective. For example, the extent to which one must disambiguate (or control for other variables) is likely to depend on how noisy, complex or densely connected the environment is. For more than about 5 or 6 variables, explicit calculation of expectancies becomes intractable while the calculations required by the active causal learning heuristics remain trivial. In every day 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 "connecting" simple endorsement and "pruning" disambiguation steps.

The interplay between forgetting and conservatism parameters in our bounded models hinted at a richness to sequential active learning which cannot be fully explained on the basis of these latent variables. In future work we would like to test peoples' memory for past trials explicitly to establish how much they forget, and whether what they remember is biased by their beliefs. We would also like to come up with a paradigm which allows us to elicit which, and how many, hypotheses people consider when choosing an intervention. As a final point, in this paradigm we removed temporal information entirely. Arguably though, interventions can be thought of as particularly strong and reliable temporal cues (Lagnado and Sloman, 2004, 0). We would like to test probabilistic causal learning in an environment in which causes take time to cause their effects, making each new outcome dependent on the state of the system at the previous time step, plus any desired intervention (Rottman and Keil, 2012). This approach could allow us to unify temporal and interventional cues to causation within a single framework.

4.2 Conclusions

In this paper we asked how people learn about causal structure through intervention. We found that many participants were highly effective active causal learners, able to select informative interventions and use these to improve their causal models incrementally over multiple learning instances. A significant percentage of participants performed at ceiling despite the task being considerably more demanding than most previous active causal learning tasks, suggesting that peoples' ability to learn actively about causal structure may have been underestimated in the past. We found that most people acted like scholars, choosing interventions likely to reduce their uncertainty about the correct structure rather than to increase their probability of being correct or expected utility. By adding forgetfulness and conservatism parameters we formulated bounded versions of these models which suggested that memory (or concentration) limitations were behind the dramatic differences in the performance of different participants, but that poor memory could be mitigated to some extent by being appropriately conservative, pre-

ferring models consistent with earlier stated beliefs. Finally we identified two simple procedures, simple endorsement and disambiguation, which provide a good fit to participants' actions and are candidate components of a heuristic model of active causal learning.

References

Baddeley, A. (1992). Working memory. *Science*, 255(5044):556–559.

Barber, D. (2012). *Bayesian reasoning and machine learning*. Cambridge University Press.

Bruner, J. S., Jolly, A., and Sylva, K. (1976). *Play: Its role in development and evolution*. Penguin.

Cartwright, N. (1989). Nature's capacities and their measurement.

Chater, N. and Oaksford, M. (2008). *The probabilistic mind: Prospects for Bayesian cognitive science*. Oxford University Press, USA.

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

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and brain sciences*, 24(1):87–114.

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

Eberhardt, F., Glymour, C., and Scheines, R. (2012). On the number of experiments sufficient and in the worst case necessary to identify all causal relations among n variables. In *UAI*. AUAI Press.

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

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

Goodman, N. D., Ullman, T. D., and Tenenbaum, J. B. (2011). Learning a theory of causality. *Psychological Review*, 118(1):110–9.

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

Gureckis, T. and Markant, D. (2009). Active learning strategies in a spatial concept learning game. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*.

Gureckis, T. and Markant, D. (2012). Self-Directed Learning: A Cognitive and Computational Perspective. *Perspectives on Psychological Science*, 7(5):464–481.

Harman, G. (1986). Change in view: Principles ofreasoning.

Heckerman, D., Geiger, D., and Chickering, D. M. (1995). Learning bayesian networks: The combination of knowledge and statistical data. *Machine learning*, 20(3):197–243.

Jones, M. and Love, B. C. (2011). Bayesian fundamentalism or enlightenment? *On the explanatory status and theoretical contributions of Bayesian models of cognition*, 34:169–188.

Kruschke, J. K. (2001). Toward a unified model of attention in associative learning. *Journal of Mathematical Psychology*, 45(6):812–863.

Kuhn, D. and Dean, D. (2005). Is developing scientific thinking all about learning to control variables? *Psychological Science*, 16(11):866–870.

Lagnado, D. and Sloman, S. (2002). Learning causal structure. In Gray, W. and Schunn, C. D., editors, *Proceedings of the twenty-fourth annual conference of the cognitive science society, Mahwah,*. Erlbaum, NJ.

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

Lagnado, D. and Sloman, S. (2006). Time as a guide to cause. *Cognition*, 32(3):451–460.

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

Markant, D. and Gureckis, T. (2012). Does the utility of information influence sampling behavior? *Proceedings of the 34th Annual Conference of the Cognitive Science Society*.

Markant, D. and Gureckis, T. (2013). Is it better to select or to receive? learning via active and passive hypothesis testing. *Journal of Experimental Psychology: General*.

Markant, D., Gureckis, T., Meder, B., Nelson, J. D., Pirolli, P., and Yu, C. (2013). Informavores: Active information foraging and human cognition. In *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*.

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

Meder, B. and Nelson, J. D. (2012). Information search with situation-specific reward functions. *Judgment and Decision Making*, 7(2):119–148.

Meier, K. M. and Blair, M. R. (2012). Waiting and weighting: Information sampling is a balance between efficiency and error-reduction. *Cognition*, 126(2):319 – 325.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological review*, 63(2):81.

Miller, G. A. (1984). Informavores. *The study of information: Interdisciplinary messages*, pages 111–113.

Murphy, K. P. (2001). Active learning of causal bayes net structure. Technical report, UC Berkeley.

Nelson, J. D. (2005). Finding useful questions: On bayesian diagnosticity, probability, impact, and information gain. *Psychological review*, 112(4).

Nelson, J. D., McKenzie, C. R., Cottrell, G. W., and Sejnowski, T. J. (2010). Experience matters information acquisition optimizes probability gain. *Psychological science*, 21(7):960–969.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, San Francisco, CA.

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

Piaget, J. and Valsiner, J. (1930). *The child's conception of physical causality*. Transaction Pub.

Puterman, M. L. (2009). *Markov decision processes: discrete stochastic dynamic programming*, volume 414. Wiley.

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

Schulz, L. (2001). Do-calculus?: Adults and preschoolers infer causal structure from patterns of outcomes following interventions. In *Second Biennial Meeting of the Cognitive Development Society*.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*,

6(2):461–464.

Shanks, D. R. (1995). Is human learning rational? *The Quarterly Journal of Experimental Psychology*, 48(2):257–279.

Shannon, C. E. (1951). Prediction and entropy of printed english. *The Bell System Technical Journal*, 30:50–64.

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

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

Sobel, D. M. (2003). Watch it, do it, or watch it done. *Manuscript submitted for publication*.

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

Sprites, P., Glymour, C., and Scheines, R. (1993). *Causation, Prediction, and Search (Springer Lecture Notes in Statistics)*. MIT Press. (2nd edition, 2000), Cambridge, MA.

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

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

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., and Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, 331(6022):1279–1285.

Tong, S. and Koller, D. (2001). Active learning for structure in bayesian networks. In *International joint conference on artificial intelligence*, volume 17, pages 863–869. Lawrence Erlbaum Associates Ltd.

Waldmann, M. (2000). Competition among causes but not effects in predictive and diagnostic learning. *Jouranl of Experimental Psychology: Learning, Memory & Cognition*, 26:53–76.

Waldmann, M. and Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: Asymmetries in cue competition. *Journal of Experimental Psychology: General*, 121(2):222–236.

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