4-2014

Discovering hidden causes using statistical evidence

Christopher G. Lucas
University of Edinburgh

Kenneth Holstein
University of Pittsburgh

Charles Kemp
Carnegie Mellon University, ckemp@cmu.edu

Published In
Proceedings of the 36th Annual Conference of the Cognitive Science Society.

Discovering hidden causes using statistical evidence

Christopher G. Lucas
c.lucas@ed.ac.uk
School of Informatics
University of Edinburgh

Kenneth Holstein
kjh64@pitt.edu
Department of Psychology
University of Pittsburgh

Charles Kemp
ckemp@cmu.edu
Department of Psychology
Carnegie Mellon University

Abstract

People frequently reason about causal relationships and variables that cannot be directly observed. This paper presents results from an experiment in which participants used statistical information to make judgments about the number and base rates of hidden causes, as well as the forms of causal relationships in which those causes participated. Our data allow us to evaluate several models of hidden cause discovery, and reveal that people have different expectations about the forms of causal relationships than recent theories predict.

Keywords: causal inference; hidden causes; minimum description length; functional form

In order to reason fluently about cause and effect, humans must frequently make inferences about causes that are hidden from direct observation. For example, we explain the behaviors of agents by appealing to mental states, we infer the presence of illnesses from symptoms, and we can diagnose hidden faults in mechanical and electrical systems by seeing how they fail. This ability is essential for explanation, prediction, and planning, but exactly how we reason about hidden causes, or even come to believe that hidden causes are present in the first place, is not yet well understood.

What are the signs that a hidden cause might be present in a causal system? One is unreliability: hidden factors are likely to be at work when a relationship appears to be unreliable or stochastic, and there is no observable reason for this behavior. Previous work has explored how and when people posit hidden causes to explain unreliable relationships. When children observe a failure-prone electromechanical device, they tend to infer that its failures are due to a hidden controller (Schulz & Sommerville, 2006). People can also infer common hidden causes when unreliable effects tend to be correlated (Kushnir, Gopnik, Lucas, & Schulz, 2010; Carroll & Cheng, 2009; Mayrhofer, Goodman, Waldmann, & Tenenbaum, 2008), and can make inferences about the nature of hidden causes from their interactions with background rates of effects (Carroll & Cheng, 2010).

Despite these results, there are many open questions about how humans use statistical information to discover hidden causes. In particular, there are no comprehensive studies that explore how people infer the form of the function that specifies how observed and hidden causes combine to produce an effect. Many theories assume that prospective causes combine according to noisy-OR and noisy-AND-NOT functions, but we introduce an experimental paradigm that reveals that people naturally entertain a broader range of causal functions.

The next section introduces the simple problem of causal discovery that we consider. The following section describes three models that capture different theories about how people might solve this problem, including a model that focuses on noisy-OR and noisy-AND-NOT relationships. We then present an experiment that allows us to compare the models and reveals patterns of judgments that standard accounts of causal reasoning do not anticipate.

A simple causal discovery problem

This paper will focus on inferences about a causal system that has a single visible cause v and a single effect e. In addition to the visible cause, there may be one or more hidden causes h_i that influence the effect. For simplicity, we assume that all variables are binary. Suppose that the system is observed over many trials – in each case the visible cause v takes value 0 or 1, and the effect e takes value 0 or 1. The resulting sequence of observations can be summarized by a pair of statistics $[p', p] = [P(e=1|v=0), P(e=1|v=1)]$ that indicates the probability that $e=1$ when v is either 0 or 1.

For our purposes, the space of all possible pairs $[p', p]$ can be organized into a handful of qualitatively different classes. Figure 1a shows representatives of the five classes that we consider in this paper. The classes can be enumerated by taking three criteria into account. First, probabilistic relationships between v and e provide evidence of hidden causes, and we therefore distinguish between extreme statistics that equal 0 or 1 and intermediate statistics that fall between 0 and 1. For example, pairs $[0,1]$ and $[p,1]$ in Figure 1a are qualitatively different because $[0,1]$ has two extreme statistics and $[p,1]$ has only one extreme statistic. The second and third criteria are based on non-generic relationships that may hold between p and p'. The second criterion requires that we distinguish between pairs for which $p = p'$ and pairs for which the statistics are not equal. The third criterion requires that we distinguish between pairs for which $p = 1 - p'$ and pairs that do not have this property. For example, pairs $[p, p]$ and $[p_1, p_2]$ in Figure 1a can be distinguished using the second criterion, and pairs $[p, 1-p]$ and $[p_1, p_2]$ can be distinguished using the third criterion. Applying the three criteria and collapsing across all other differences between pairs produces a total of six equivalence classes. The class not shown in Figure 1a includes the pairs $[1,1]$ and $[0,0]$, and we dropped this class because the effect variable is constant.

Although the relationship between v and e may appear to be probabilistic, we assume that e is a deterministic function of v and potentially one or more hidden variables. This assumption of causal determinism is consistent with several previous formal approaches (Pearl, 2000; Lucas & Kemp, 2012) and with previous suggestions that people are causal determinists (Schulz & Sommerville, 2006; Frosch & Johnson-Laird,

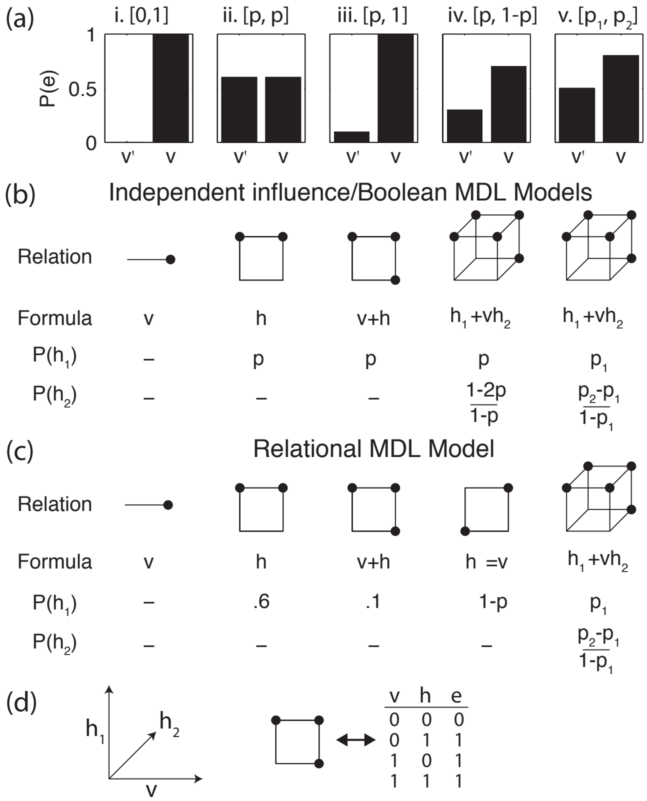

Figure 1: Predictions of the three models of hidden cause discovery. (a) Examples of the different classes of statistics that are possible in a single-cause scenario. (b) Predictions of the independent-influence and Boolean MDL models. (c) Predictions of the relational MDL model. The $P(h_1)$ and $P(h_2)$ values assume, without loss of generality, that $p < (1-p)$ in the $[p, 1-p]$ case, and $p_1 < p_2$ in the $[p_1, p_2]$ case. (d) Explanation of the box diagrams in (b) and (c). The horizontal dimension of each box represents variable v, and additional dimensions are shown when hidden variables h_1 and h_2 are postulated. Black dots show settings of the variables that produce the effect. For example, the box diagram shown corresponds to a truth table for disjunction.

2011). Given this assumption, a complete causal explanation of a system characterized by the pair $[p', p]$ must provide answers to three questions. The first asks whether the system includes any hidden variables, and if so how many. The second asks about the base rates of any postulated hidden variables. The third asks about the functional form of the system, and must be answered by providing a function that allows e to be predicted with certainty given the values of v and any hidden variables.

Theories of causal discovery

We now describe three models that represent different perspectives on how people might solve the problem just introduced. All three propose that people tend to generate the simplest explanation of the data, and introduce hidden causes as needed so that the function that determines the value of effect e can be expressed as simply as possible. The key difference between the models is that they make different assumptions about the forms that causal relationships can take.

Independent influence model. A common foundation of many accounts of causal reasoning is the assumption of *causal independence*. This assumption requires that each cause influences the effect in a way that is independent of the value of other causes. The independence requirement is typically satisfied by assuming that generative causes combine to produce an effect according to a noisy-OR function, and that preventive causes combine according to a noisy-AND-NOT function. Our first model therefore assumes that the value of effect e is determined by the formula

$$(c_1 f_1' + \ldots + c_n f_n')(c_{n+1}' + f_{n+1}) \ldots (c_{n+m}' + f_{n+m}) \qquad (1)$$

where c_1 through c_n are generative causes, c_{n+1} through c_{n+m} are preventive causes, and each cause c_i is associated with a failure variable f_i that determines whether the mechanism linking c_i to e fails on a given trial. Negations are represented using primes, and conjunctions and disjunctions are represented using products and sums respectively. For example, $c_1 f_1'$ is true if cause c_1 is present AND the failure variable f_1 is false. Similarly, $c_{n+1}' + f_{n+1}$ is true if cause c_{n+1} is absent OR f_{n+1} is true.

The independent influence model proposes that people explain the observed statistics in terms of the simplest function that matches the schema in Equation 1. For example, if a causal system is characterized by the pair $[p, 1]$, then the effect sometimes occurs when v is absent, and an additional generative cause h with base rate p is inferred such that that e equals $v+h$. In this case, the failure variables for v and h both have base rates of zero, and we therefore allow them to be dropped. If the pair for the system is $[p_1, p_2]$, then there must be at least two hidden variables, but these variables could combine in different ways. For example, h_1 could be a generative cause and h_2 could be the failure variable for v, which would yield the formula $h_1 + vh_2'$. Alternatively, h_1 could be a generative cause and h_2 could be a preventive cause, which would yield $(h_1 + v)h_2'$. Both explanations are viable according to the independent influence model, because both formulas include three variables and therefore have a complexity of 3.

Predictions of the independent influence model for five pairs of statistics are shown in Figure 1b. The formulas that determine e are shown along with the base rates of any postulated hidden variables. For each of the five cases only a single minimal explanation is shown. This minimal explanation is one instance of a class of equivalent explanations: for example, $v+h$ in the third column can be replaced by $v+h'$ if the base rate of h is set to $1-p$ rather than p.

Several previous researchers have pointed out that the assumption of causal independence supports the discovery of hidden causes, and our independent influence model is broadly consistent with the work of Carroll and Cheng (2010) and Mayrhofer et al. (2008). We have not attempted to characterize the process by which a minimal explanation is identified, but one possibility is that people postulate hidden causes incrementally, introducing each one only when necessary to explain effects that depend in a stochastic way on the variables observed or postulated thus far.

Boolean MDL model. The independent influence model can be viewed as a *minimum description length* (MDL) model because it infers the minimal instance of Equation 1 that accounts for the available data. Our second model shares this emphasis on minimal explanations, but works with a space of formulas that includes all Boolean functions, not just the functions consistent with Equation 1. For example, one explanation of the pair $[p, 1-p]$ is the formula $vh + v'h'$, where hidden cause h has a base rate of $1-p$. This formula is not permitted by the independent influence model, which assumes that any cause in the set $\{v, h_1, h_2, \ldots\}$ can only bind to a single variable in Equation 1. The idea that arbitrary Boolean formulas may be required to capture causal systems is consistent with some previous proposals (Yuille & Lu, 2008; Buchanan, Tenenbaum, & Sobel, 2010), and is also closely related to Feldman's work on concept learning in a non-causal setting (Feldman, 2000). As for the independent influence model, the Boolean MDL model defines the complexity of a formula as the number of variable tokens that it contains.

Predictions of the Boolean MDL model are shown in Figure 1b. Although the model allows a broader class of explanations than the independent influence model, the two make identical predictions for the cases shown in Figure 1. For example, even though the Boolean MDL model can entertain the formula $vh + v'h'$ for the pair $[p, 1-p]$, this formula has complexity 4 and is therefore more complex than the formula $h_1 + vh_2$.

Relational MDL model. Our third model is similar to the Boolean MDL model, but proposes that people draw from a broader repertoire of representational primitives than is captured by Boolean logic. The relational MDL model extends the Boolean MDL model by allowing formulas that capture equality and inequality relationships between causes. While these relations tend to be absent from discussions of causal learning, it seems plausible that people can make use of equality and inequality relations when reasoning about causal systems.

Predictions of the relational MDL model are shown in Figure 1c. The model makes the same predictions as the previous two models in all cases except one. The equality relation allows the $[p, 1-p]$ pair to be explained using the formula $h = v$, which indicates that e is present only if the value of h matches the value of v. This formula has complexity 2, and is therefore simpler than the simplest formula allowed by Boolean logic alone.

In the next section, we describe an experiment that compares the predictions of the three models against human judgments about the five classes of pairs shown in Figure 1a. The experiment has four possible outcomes. First, it may be that human learners find it difficult or impossible to draw coherent inferences when faced with such limited evidence and no auxiliary evidence such as temporal information. In this case,

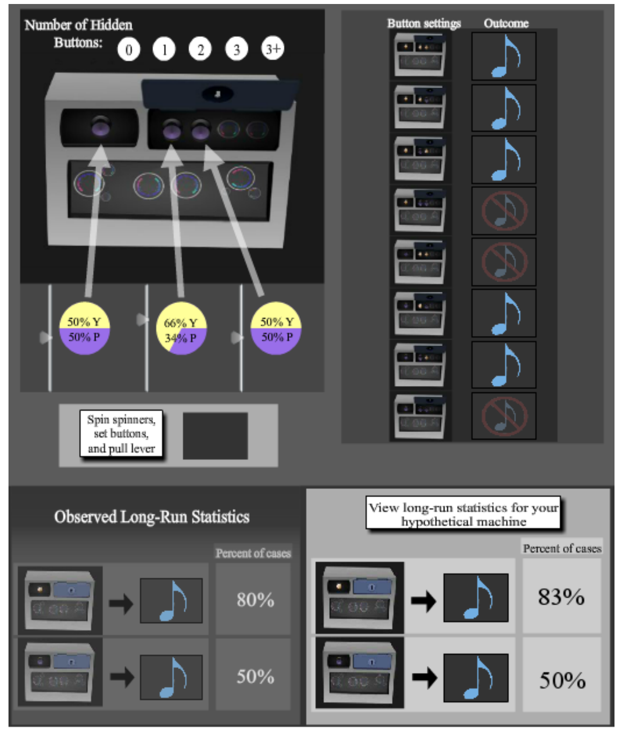

Figure 2: The interface used by participants to design their own alien machines. Adjustable spinners are displayed under an image of the alien machine. To the right of the machine is an interactive table displaying all possible button settings for this machine, and possible outcomes for each of these settings. In the bottom right, clicking the "View long-run statistics" button allows participants to view the long-run frequencies for their machine under each setting of its visible button. In the bottom left is a reminder of the long-run frequencies observed in the original machine. The machine above corresponds to the $[0.5, 0.8]$ instance of the $[p_1, p_2]$ condition in Table 1, in which a sound is produced in 80% of cases where the visible button was set to yellow and 50% of cases where the visible button was set to purple.

participants will offer relationships that are incapable of capturing the observed statistics, or will make inferences about the rates of hidden causes that are inconsistent with the relationships they posit. A second possibility is that participants make judgments that are consistent with the predictions of the independent influence and Boolean MDL models, which would support recent proposals about hidden cause discovery (Carroll & Cheng, 2010; Buchanan et al., 2010). A third possibility is that the relational MDL model accurately predicts human inferences. The fourth and final possibility is that people make coherent inferences about hidden causes that are not consistent with any of these models, implying that the representational machinery that supports causal inference (and hidden cause discovery in particular) is more complex than is often assumed.

Experiment

Our experiment used causal systems that were described as alien machines, and one such machine is shown in the top left region of Figure 2. Each machine had a single visible button with two possible settings, yellow and purple. Each machine also had a panel that could potentially conceal additional, hidden buttons. The panel for the machine in Figure 2 is open, revealing two hidden buttons. Each button had an associated "spinner" that was used to determine the setting of the button. These spinners were circular disks resembling pie charts, which could be colored yellow and purple in any proportions. Three examples are shown in Figure 2.

The aliens were said to use these machines by repeatedly carrying out a sequence of actions known as a Spin-Set-Pull. A single Spin-Set-Pull involves spinning all of the spinners, setting the buttons accordingly, then pulling a lever (not shown in Figure 2). Each Spin-Set-Pull either causes or does not cause the machine to produce a sound.

Our experiment presented each participant with five conditions corresponding to the families in Figure 1a. In each case, participants observed the pair of statistics associated with a given machine, but did not observe how many buttons (if any) lay behind the panel and did not observe the spinners associated with any of the buttons. Observations for one condition are shown in the bottom left of Figure 2 under the heading "Observed Long-Run statistics." On the basis of these observations, participants were asked to report the number of hidden buttons, the base rates (i.e. the spinner proportions) for all buttons, and the function that determined the effect (sound or no sound) given the settings of the buttons.

Methods

Participants. Fifty-four Carnegie Mellon University undergraduates participated in the experiment for course credit.

Materials and Procedure. Participants were provided with a cover story that described the alien machines. To help participants understand the scenario and their task, they were also given a brief tutorial. Participants were first shown an example machine and asked to perform a Spin-Set-Pull on this machine. They then learned how to use an interface that allowed them to design their own alien machines (see Figure 2). This interface allowed participants to select the number of hidden buttons their machine would have, with options ranging from 0 to 3 hidden buttons, and an additional option to specify a machine with more than 3 hidden buttons.

When a participant selected between 0 and 2 hidden buttons, an open-panel machine with the corresponding number of hidden buttons and spinners was displayed, along with a table showing the outcomes tied to each of the possible button settings for that machine. For example, Figure 2 shows a case in which a participant has specified a machine with 2 hidden buttons, and as a result a machine with 3 buttons and 3 spinners is displayed, alongside a table with one row for each of this machine's 8 possible button settings. Participants could then interact with the table to specify whether or not the

Table 1: Instances of the 5 classes presented to participants.

Condition	Instances
$[0,1]$	$[0.0,1.0]$, $[1.0,0.0]$
$[p,p]$	$[0.6,0.6]$, $[0.4,0.4]$
$[p,1]$	$[0.1,1.0]$, $[1.0,0.1]$, $[0.0,0.9]$, $[0.9,0.0]$
$[p,1-p]$	$[0.3,0.7]$, $[0.7,0.3]$
$[p_1,p_2]$	$[0.5,0.8]$, $[0.8,0.5]$, $[0.2,0.5]$, $[0.5,0.2]$

machine would produce a sound under each possible setting, and could also adjust the proportions of yellow and purple on each button's spinner. After fully specifying a machine, participants could click a "View long-run statistics" button to view the long-run frequency with which their machine would produce a sound, under each setting of the visible button. To help participants understand the relationship between these long-run frequencies and individual Spin-Set-Pulls, participants could also click a button to view single Spin-Set-Pulls and their outcomes.

After learning to use the interface, participants worked through 5 within-subjects conditions presented in random order. Participants were told that each condition represented a different alien machine, each with a closed panel and spinner cabinet. In each condition, participants were told the long-run frequencies with which an alien machine produced a sound under each setting of the visible button. Using this information, participants were asked to use the interface to show how they thought the machine worked. Participants were told that they could experiment with different designs for their machine, but that the final state of the interface should reflect their best guess about how the original machine worked. Each time a participant clicked on the "View long-run statistics" button, the interface recorded the participant's current machine design. Once participants were finished using the interface in a given condition, they were asked to specify whether or not they believed that their machine produced long-run frequencies that exactly matched those of the original machine, and also to give a free-text response that characterized the button settings that caused their machine to produce a sound. If a participant specified a machine with 3 or more hidden buttons, the interface did not display spinners or a table of outcomes, and the participant was instead asked to respond only to these last two questions.

The 5 conditions presented long-run frequencies that corresponded to the 5 classes shown in Figure 1a. Several different instances of each class were used, as shown in Table 1, and the instances used for each participant were chosen randomly from these options.

Results

Participants provided three kinds of information about each alien machine: the number of hidden buttons they believed to be present, the base rates for each button, and the function that determined the value of the effect given the settings of all buttons. Participants could interact with each machine

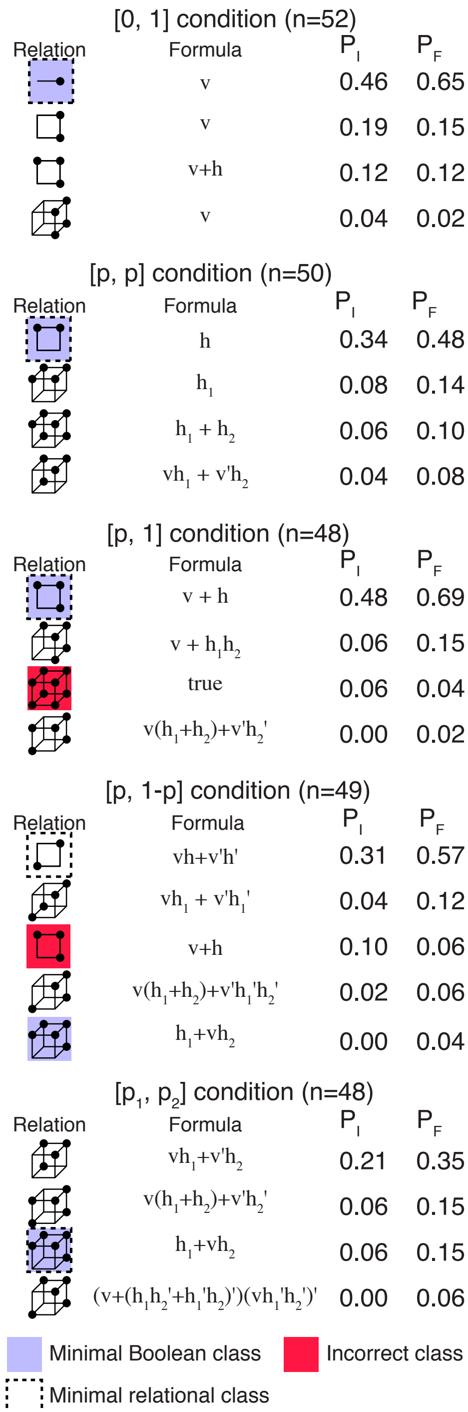

Figure 3: Judgments about causal relationships. Column 1 represents the most common relationship classes. Backgrounds mark the simplest classes according to the Boolean MDL and relational models, as well as incorrect relationships that cannot explain the observed statistics. See the Results section for details on the classes and notation. When the simplest relationship was not among the four most common guesses, it was appended as a fifth row. The second column contains the corresponding minimal Boolean formula. Columns 3 and 4 show proportions of initial (P_I) and final judgments (P_F) judgments in each class, sorted by P_F.

as many times as they wanted, opening the possibility that their final judgments reflect the results of lengthy searches that involve repeated guessing and checking. To address this issue, we also report participants' initial judgments: that is, the judgments that they made about a given machine before clicking the "View long-run statistics" button for the first time. The patterns of initial and final judgments were very similar, as shown in Figures 3 and 4.

There were a total of 54 participants who saw five conditions each, but data were only available for 247 of the 270 possible (participant,condition) pairs due to server errors (2 pairs) and participants who inferred more than 2 hidden buttons or never used the "View long-run statistics" option when specifying a machine (21 pairs). Figure 4 summarizes participants' judgments about the numbers of hidden causes. Figure 4 shows that for every condition, the modal number of hidden causes in participants' final judgments corresponded to the minimal number necessary to produce the observed statistics[1]. Excluding $[p, p]$, the minimum was chosen significantly more often than half of the time (binomial tests, $p < .05$). For participants' initial judgments, the same pattern held in every condition but $[p, p]$. Figure 4 therefore suggests that participants were inclined to introduce hidden causes only when required to by the observed statistics.

Figure 3 summarizes participants' judgments about the functional form of each system. The formula column in Figure 3 gives the shortest formula that describes the relationship, under the Boolean MDL model. For the alien machines, x and x' denote different values of the variable x, i.e., purple or yellow states for the three buttons. The formula notation is described in the Theories section. In order to visualize our data without losing the ability to address our main questions, we combined the distinct instances of each condition and merged judgments into equivalence classes as follows. First, we treated the hidden buttons as exchangeable, so that, for instance, the relationship described by $h_2 + vh_1$ is in the same class as that described by $h_1 + vh_2$. Second, we treated the different values of the effect (sound or no sound) as being equivalent, so that $h + v$ and $(h + v)'$ are in the same class. Third, we treated values of the causes as being equivalent, so that $h_1 + h_2'$ and $h_1 + h_2$ are in the same class.

Figure 1 shows that all three models agree on the best explanations for the $[0, 1]$, $[p, p]$ and $[p, 1]$ conditions, and Figure 3 shows that the same explanations were the most common human responses for these conditions. The models make different predictions for the $[p, 1 - p]$ condition, and the data in Figure 3 are more consistent with the relational MDL model than the other two models. Even though many accounts of causal reasoning rely on noisy-OR/noisy-AND-NOT parameterizations, our participants seemed willing to invoke alternative functional forms that are sensitive to whether two variables take equal values.

[1] This minimum number was determined by assuming that each hidden button's state was determined independently, and finding the minimum number for which it was possible to find base rates that perfectly reproduced the observed effect rates.

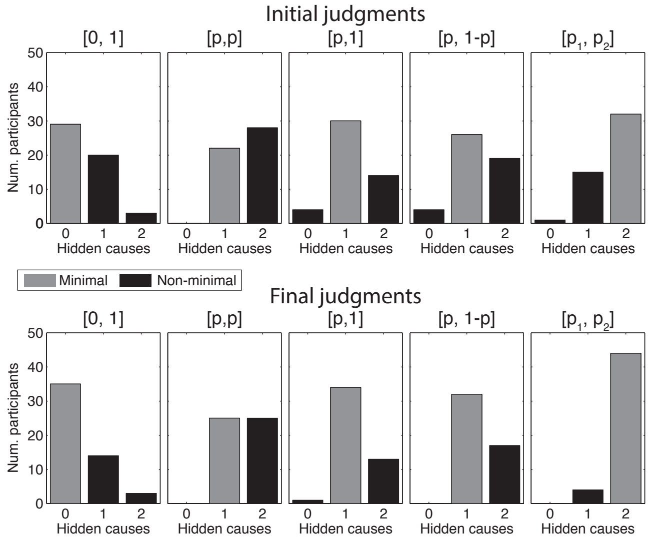

Figure 4: Numbers of individuals who judged 0, 1, or 2 hidden causes to be present, grouped by condition. The gray bars correspond to the minimal number of hidden causes that are necessary to explain the observed statistics.

Although the relational MDL model is broadly consistent with the responses to the first four conditions, it does not predict the most common response to the $[p_1, p_2]$ condition. All three models predict that the standard noisy-OR parameterization $h_1 + vh_2$ should be the preferred explanation for this condition. Figure 3 shows that this response was given by some participants, but was dominated by $vh_1 + v'h_2$, which specifies a function in which the visible button plays the role of a switch: if v is purple then the value of e is determined entirely by h_1, and if v is yellow then the value of e is determined entirely by h_2. One reason why this "switch" explanation may have been attractive is that the base rates for the two hidden variables are identical to the long-run statistics p_1 and p_2 that were provided as part of the description of the machine. In contrast, the noisy-OR explanation means that some thought is required to determine the base rate for one of the variables: in Figure 1, for example, the base rate of h_2 for the $[p_1, p_2]$ condition is a function of both p_1 and p_2. Additional work will be necessary to determine whether this preference for the switch explanation is an artifact of participants being asked to provide numerical probabilities, or a more general phenomenon.

Judgments about functional forms and the rates of hidden causes determine the long-run statistics for a hypothesized machine. We can compare these hypothesized statistics to the long-run statistics that participants actually observed to assess their accuracy in explaining the behavior of the systems. For both first and last judgments, their mean squared error rates ($M = .15$, $SD = .19$ and $M = .10$, $SD = .19$, respectively) were significantly lower than what would have been observed if their judgments had not been sensitive to condition ($p < .001$ in both cases; permutation tests). Of the 54 participants, 5 had initial judgment mean squared errors of less than .01. For final judgments, 10 participants had mean square errors of less than .01. These analyses provide additional evidence that many participants were indeed able to make coherent inferences about hidden causes on the basis of the limited information provided.

Conclusion

Our results demonstrate that humans can not only infer the presence of hidden causes from statistical evidence, but also make coherent inferences about how hidden and observed causes combine to produce an effect. Many theories of causal learning assume that prospective causes influence an effect independently, but our data suggest that people readily violate this assumption in favor of explanations that invoke relationships between prospective causes. Characterizing the full range of basic elements that people use to construct causal explanations is an important goal for future work.

Our analysis combined different kinds of causal relationships into a small number of equivalence classes. This approach provided a simple way to evaluate the predictions of different kinds of models, but our experimental paradigm can also be applied to questions that demand finer distinctions between relationships. For example, pairs $[0, 0.9]$ and $[0.1, 1.0]$ were both treated as instances of class $[p, 1]$, but the first pair corresponds to a conjunction vh and the second pair corresponds to a disjunction $v + h$. Whether people find conjunctive relationships easier to discover than disjunctive relationships is just one of many additional empirical questions that may repay investigation.

References

Buchanan, D., Tenenbaum, J., & Sobel, D. (2010). Edge replacement and nonindependence in causation. In *Proceedings of the 32nd Annual Conference of the Cognitive Science Society.*

Carroll, C. D., & Cheng, P. (2009). Preventative scope in causal inference. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 833–838).

Carroll, C. D., & Cheng, P. W. (2010). The induction of hidden causes: Causal mediation and violations of independent causal influence. In *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 913–918).

Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature*, *407*, 630-633.

Frosch, C. A., & Johnson-Laird, P. (2011). Is everyday causation deterministic or probabilistic? *Acta psychologica*, *137*(3), 280–291.

Kushnir, T., Gopnik, A., Lucas, C., & Schulz, L. (2010). Inferring Hidden Causal Structure. *Cognitive Science*, *34*(1), 148–160.

Lucas, C. G., & Kemp, C. (2012). A unified theory of counterfactual reasoning. In *Proceedings of the 34th Annual Meeting of the Cognitive Science Society* (pp. 707–712).

Mayrhofer, R., Goodman, N. D., Waldmann, M. R., & Tenenbaum, J. B. (2008). Structured correlation from the causal background. In *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 303–308).

Pearl, J. (2000). *Causality: Models, reasoning and inference*. Cambridge, UK: Cambridge University Press.

Schulz, L., & Sommerville, J. (2006). God does not play dice: Causal determinism and preschoolers' causal inferences. *Child Development*, *77*(2), 427-442.

Yuille, A. L., & Lu, H. (2008). The Noisy-Logical Distribution and its Application to Causal Inference. In J. Platt, D. Koller, Y. Singer, & S. Roweis (Eds.), *Advances in Neural Information Processing Systems 20* (pp. 1673–1680). Cambridge, MA: MIT Press.