# Edge replacement and nonindepedence in causation

**David W. Buchanan (david_buchanan@brown.edu)**
Department of Cognitive and Linguistic Sciences
Box 1978, Brown University, Providence, RI, 02912

**Joshua B. Tenenbaum (jbt@mit.edu)**
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology

**David M. Sobel (dave_sobel@brown.edu)**
Department of Cognitive and Linguistic Sciences
Box 1978, Brown University, Providence, RI, 02912

**Abstract**

Human beings show a robust nonindependence effect in causal reasoning: they predict that collateral effects should be correlated even given a common cause. This presents a problem for existing models of causal reasoning, as most predict independence. To deal with this problem, we propose an edge replacement process that builds up apparently probabilistic causal relations using hidden deterministic causes. This model allows us to fit nonindependence effects, and shows promise for modeling other phenomena, such as how causal relations change over time.

**Keywords:** Markov violations; nonindependence; causal reasoning; models of causal reasoning

## Introduction

Causation is only as simple as we make it. Consider the example of sending an email to two colleagues: You push send, which causes them to see text on their screen. The relation seems simple enough, but in reality, there is a complex chain of events that connects cause and effect, which most of us understand only vaguely (Keil, 2003). These details are usually not worth considering, but they are useful when the causal relations fail. For instance, most of us know to check our spam filter when we fail to receive an expected email. Such details also tell you about correlations between events: If one colleague calls to say she has not received the message, you know to call the other one as well. Still, more detail is not always better – it would be absurd reason about email at a molecular level. Choosing the right level of detail is important, and human beings seem to do it easily. Models of causal inference must solve this problem as well.

Causal graphical models (hereafter, "CGMs"), (Pearl, 2000; Spirtes, Glymour, & Scheines, 2000) give us a language in which to express this problem formally. Under this framework, nodes in a graph represent events, and directed edges represent causal relations. Figure 1 gives an example of three graphs that capture the common cause scenario described above: person $C$ sends an email, causing persons $E_1$ and $E_2$ to receive it. Under the assumptions of CGMs, unconnected nodes must be statistically independent, but otherwise there are a wide range of possible functional relationships that can be instantiated by an edge. There is also no limit to the number of hidden nodes that can exist in a graph.

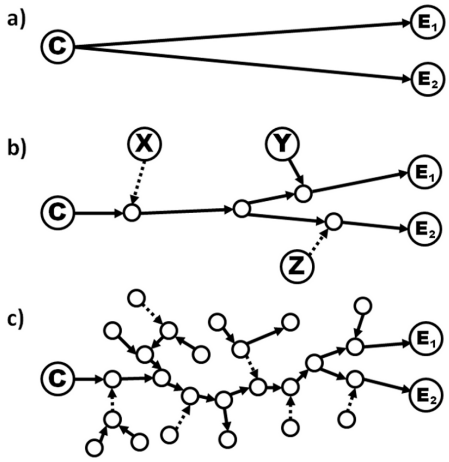


Figure 1: Examples of three different graphs that all capture the common cause relation. Minimality prescribes that we should begin by using a).

Thus, CGMs are enormously powerful, defining an infinite space of possible graphs for any given causal relation.

In order to make use of CGMs, we need some way of choosing which graph to use. The principle currently most used is called *minimality*: use the simplest graph that fits the data, in the sense that no other candidate graph has fewer edges. This means that given a common cause of two effects, the minimal graph is shown in Figure1a. It is often acknowledged (i.e. Pearl, 2000) that minimality creates problems, but in the absence of an alternative, it is widely used.

Within the literature on causal reasoning, the most acute problem with minimality is known as *nonindependence*. The simplest example of this phenomenon is found in a common cause scenario. If two effects of a common cause are related according to the minimal graph (Figure 1a) then the two effects should be independent given their common cause. That is, if $C$ directly causes each of $E_1$ and $E_2$, then $P(E_1|C,\neg E_2)$[1]$= P(E_1|C) = P(E_1|C,E_2)$. If we see evidence that violates this expectation, then minimality allows us to use

[1]This notation means: The probability that $E_1$ occurs given that $C$ occurred, but $E_2$ did not.

a slightly more complex graph that better fits the data. But according to minimality, independence should be our initial expectation.

Human beings do not have this expectation. In several experiments (Mayrhofer, Hagmayer, & Waldmann, 2008; Rehder & Burnett, 2005; Walsh & Sloman, 2004) participants robustly predict that $P(C|E_1, \neg E_2) < P(E_1|C) < P(E_1|C, E_2)$, even in novel scenarios, and even when independence is explicitly emphasized. Such nonindependence effects show that if people respect CGMs, they do not respect minimality. This raises the question of what principle, if any, people do respect.

Rehder and Burnett (2005) proposed an "underlying mechanism model" to address this problem, in which people represent hidden intermediate causal structure. In its current form, this model allows only qualitative fits to the data. Our model is one way of formalizing and extending Rehder and Burnett's proposal in order to make quantitative predictions. Further, Mayrhofer et al. (2008) modeled nonindependence effects using a source of common preventative noise, whose strength they fit to the data. Again, we hope to build on this initial step, and account formally for the source of this noise in a more principled way, while using fewer parameters to fit experimental data.

We propose a generative model of causation, which we call the *causal edge replacement process* (CERP). Theoretically, it is motivated by the hypothesis that causal reasoning involves representations of intermediate causal structure, or mechanisms (Shultz, 1982). Formally, CERP assigns a probability distribution to an infinite space of possible graphs, depending on how likely each is to be generated using repeated application of a specific edge replacement rule, and a restricted function set. The model's key contribution is that in the generative process, each edge has a *length*; longer edges tend to generate more hidden structure. While the graphs preferred by the model tend to be simple, they are not minimal. In particular, graphs generated by CERP have a characteristic branching structure that gives good quantitative fits to human data on nonindependence. CERP also provides a formal way of addressing questions about causal mechanisms.

We will begin by explaining exactly how the generative model operates, then show how the model fits three independently collected data sets, using the same parameter settings. Finally, we will discuss directions for future work.

## Generative model

We will describe the generative model in three independent ways: In this section, we will give an informal verbal overview of the edge replacement process. Figure 2 also shows the process visually. Finally, we will describe the model in complete formal detail.

CERP begins with an edge between two nodes, which represents a causal relation between two events of interest. The process then moves down this edge, randomly generating replacements as it goes. Each replacement incorporates the influence of a new node. Because of the branching structure created by CERP, it is helpful to think of causation as flowing down a stream from causes to effects.

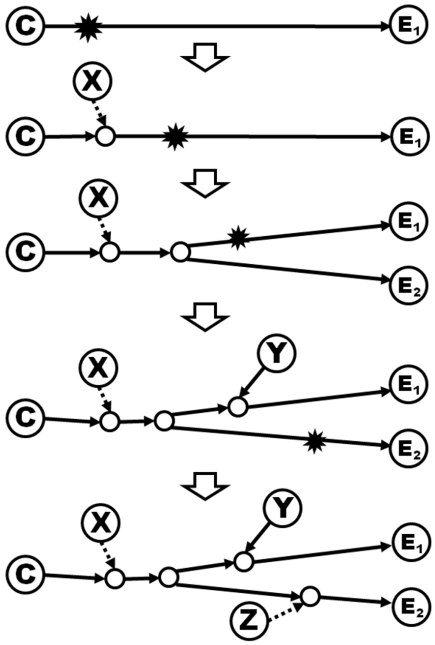


Figure 2: A series of edge replacements leading to a graph. The star indicates the location of the next replacement, as the process moves down each edge. Dashed edges indicate inhibitory relations.

To perform a replacement, we first replace the edge with an edge-node-edge path combination that has the same length. We call this middle node the "bridge node." Then we add a new edge (of randomly determined length) connecting the bridge node and a new node. The meaning of "length" here is more functional than physical – it does not correspond to the spatial distance between cause and effect, but to how vulnerable the relation is to other events. The replacement is randomly determined to be one of three types: If it is an inhibitory replacement, then causation at this point follows an AND NOT relation: causal power will flow through the bridge node only if the new node is off. For instance, in Figure 2, $X$ is generated by an inhibitory replacement. This might be a failure in the server that sends your email. (These specific cover stories are not generated by CERP; they are only used to illustrate the principles involved.) Replacements can also be generative: For instance, $E_2$ is generated next – it fires whenever causation reaches the bridge node. We call these "side effects": For instance, sending an email may leave a record on the server. The third type of replacement is also generative, but causation flows inward rather than outward. We call these "alternative generative causes," because they follow an OR relation: the effect fires either if causation reaches the bridge node, or if the alternative generative cause fires. For instance, see $Y$ in Figure 2: this might repre-

sent the fact that you can cause a given colleague to see the text on their screen via another method, like directing them to a website. In principle, outward inhibitory replacements are also possible, but are usually irrelevant – we only include them for the sake of formal completeness results discussed below. When the new node (not the bridge node) is generated, it is assigned some random probability of firing – this is the only source of randomness in the causal structures defined by CERP. Thus CERP is committed to determinism, in the weak sense that variability arises from hidden causes, not from intrinsic randomness in causal relations.

After a replacement, the same process continues along the old path, and along the new edge created by the replacement. Thus, graphs can become arbitrarily complex if replacements are common enough. The process eventually stops when it reaches the end of all edges, yielding a graph. We can "run" the graph by deciding (again, randomly) whether each exogenous node is on, then propagating causation deterministically through the graph.

## Formal Description of CERP

This section describes the model in complete formal detail. Readers who are not interested in implementing CERP can skip this section. A graph with $n$ nodes consists of the following components: 1. An $n \times n$ matrix $G$ that encodes the causal relations (edges) between nodes. (1=generative, -1=inhibitory, 0=no relation) 2. An $n \times n$ matrix $L$ of edge lengths. 3. A vector $S$ of length $n$ that encodes the spontaneous activation probabilities of each node. Together, $< G, L, S >$ defines a graph.

The generative process begins with an edge of length 1 between two nodes, which we will call $A$ and $B$. We perform replacements by moving along each edge and generating replacements according to a Poisson process with rate $\lambda$. This is done by sampling $x$ from $Exponential(1/\lambda)$. If $x > L(A,B)$, then stop. Otherwise do a replacement at point $x$: Create a new node $M$ as the $(n+1)$th node. With probability $\rho$, designate a previously generated non-bridge node as $E$, otherwise create a new node $E$ as the $(n+2)$th node (For our purposes in this paper, $\rho = 0$). Set $G(A,M) = 1$ and $G(M,B) = G(A,B)$. Set $L(A,M) = x$ and $L(M,B) = L(A,B) - x$. If $E$ already exists, and it is exogenous, set $G(E,M)$, and if it is endogenous, set $G(M,E)$. Otherwise, with equal probability, choose to set either $G(E,M)$ or $G(M,E)$. Set this relation as $-1$ or 1 with equal probability. Also sample $L(M,E)$ or $L(E,M)$ from $Exponential(\gamma)$. Set $S(n+1) = 0$ and sample $S(n+2)$ from $Beta(\alpha, \beta)$. Finally, set $G(A,B) = 0$, eliminating the original edge. Initiate two new Poisson processes, along $MB$ and $ME$, and repeat until all processes have stopped.

To sample from the graph, determine whether each node is on, according to $S$, then propagate causation deterministically through the graph to determine the values of each non-exogenous node. A node is on if and only if all of its inhibitory connections are off, and at least one of its generative connections is on, or it fires spontaneously. This instantiates the OR (for generative) and AND NOT (for inhibitory) functions originally applied to causation by Cheng (1997).

## Completeness and Validity

We can use CERP to construct any logical relation: OR can be created by an inward generative replacement, AND NOT from an inward inhibitory replacement, while AND can be created by an inhibitory replacement on the negation of a variable. In particular, the inward inhibitory replacement acts as a "causal transistor," letting us construct a wide range of logical "circuits." We can also use the presence of a hidden inhibitor to generate any apparently probabilistic relation between two variables: To generate any $P(B|A)$, perform a hidden inhibitory replacement on the edge $AB$, with spontaneous activation probability $1 - P(B|A)$. Similarly, for any $P(B|\neg A)$, perform a hidden generative replacement with spontaneous activation probability $P(B|\neg A)$.

Yuille and Lu (2008) show that their noisy-logical graphs can capture any causal-functional relation. If we additionally allow CERP to reuse existing exogenous nodes when performing replacements (i.e. $\rho > 0$), then it is easy to see that CERP can be used to mirror any noisy-logical graph, as we can construct any logical or apparently probabilistic relation as described above. Thus, we can extend Yuille and Lu's (2008) completeness result to CERP. Some such relations will be generated with low probability by CERP, but all relations will have nonzero probability of being generated. Thus, CERP defines a prior distribution over the hypothesis space of all possible causal-functional relations.

The model also preserves validity: Because it introduces no undirected edges or cycles, it will always produce a directed acyclic graph when given a directed acyclic graph. To introduce a cycle, the model would have to introduce a path from a descendant to an ancestor. But this is not possible, because all new paths are either from nodes that have no ancestors, or to nodes that have no descendants.

Overall, CERP provides a way of expressing causal-functional relations using a compact set of rules. The restriction to deterministic OR and AND NOT functions means that complex relations must be expressed graphically, where the complexity is easier to see and measure, than it is in the complex conditional probability tables often used in existing instantiations of CGMs.

The model has two key components: the idea of edge replacement, and the use of deterministic causal relations. It is conceivable that we could use edge replacement with probabilistic relations. For instance, edges could begin with probabilistic values that change as replacements are made. However, this introduces a great deal of complexity, which is unwarranted unless necessary to fit human data. Given evidence (e.g. Schulz & Sommerville, 2006) that even children seem to be deterministic in the relevant sense, we believe it is a good assumption. Future work will focus on testing this determinism commitment directly. In this paper, we will instead focus on testing the structural predictions that arise primarily from the use of length in the generative process.

## Fitting nonindependence effects

### Walsh and Sloman (2004)

Walsh and Sloman (2004) showed a nonindependence effect that provides the simplest test of our model. They told adult participants simple common effect cover stories, in which event $C$ caused both events $E_1$ and $E_2$. For instance, some participants were told that jogging ($C$) caused both increased fitness ($E_1$) and weight loss ($E_2$). They then asked participants to judge $P(E_1|C)$ and $P(E_1|C,\neg E_2)$. Figure 3 shows their data averaged across experiments. Participants reliably judged that $P(E_1|C) > P(E_1|C,\neg E_2)$. This is a nonindependence effect: If both effects are generated independently from the cause (as in Figure 1a), both values should be the same.

CERP's predictions are also shown in Figure 3. In fitting this and subsequent experiments, we used Monte Carlo sampling on causal structures as generated by CERP. One approach would be to generate a large set of graphs using CERP, keeping only the small subset that are consistent with the cover story and data presented to participants. In this case, we would accept only graphs that had exactly two visible effects. A sufficiently large sample will reflect the properties of the probability distribution defined by CERP. Such an approach is correct but computationally expensive, prohibitively so as we add complexity to the cover story.

We used a more efficient, but equivalent procedure: Begin with a single edge between the cause $C$ and effect $E_1$ described to participants. Then generate the single visible side effect described to participants. This is equally likely to have been generated from any given point on the path from cause to effect, so we generate the second effect by choosing a random point $x \sim U[0,1]$.[2] Call the branch point $M$, and set the length of $ME_1$ to $(1-x)$ in order to ensure that $E_1$ and $E_2$ have the same path length from $C$ and hence the same expected $P(E_n|C)$. This is in order to meet the condition, common in nonindependence experiments, that effects are equally likely given the cause. Because of this equivalence, the choice of initial effect in CERP is arbitrary. This process creates three edges: $CM$, $ME_1$, and $ME_2$.

At this point, we have generated all the visible causes and effects described to participants. Therefore, we are licensed in using a computational shortcut to do simultaneous inference over all the further (hidden) replacements that could be generated by CERP. In this case, all that matters are inward hidden replacements that occur along each edge. Active replacements on $CM$ (like $X$ in Figures 1 and 2) change both relations (creating a correlation); active replacements on $ME_i$ (like $Y$ and $Z$ in figures 1 and 2) change only the relation $CE_i$, and inactive replacements have no effect.

We introduce the parameter $h$ to describe a Poisson process that moves along the edge of interest, generating only active inward hidden replacements whose causal power actually reaches the path, with rate $-ln(h)$. This means that the probability of having zero active replacements along an edge of length $l$ is $h^l$. We also introduce the parameter $a$ to capture the probability that a given visible event fires spontaneously, as a result of causal processes not captured in the graph (Cheng, 1997). Together, $h$ and $a$ replace all the parameters described above in the full formal model. For instance, for most values of $h$ and $a$, there are many settings of $\lambda$, $\gamma$, $\alpha$ and $\beta$ that produce the same predictions. As long as we are not interested in the specific causal structure that generated each effect, this process is equivalent to generating a larger sample of more detailed graphs.[3] Crucially, it also uses fewer parameters.

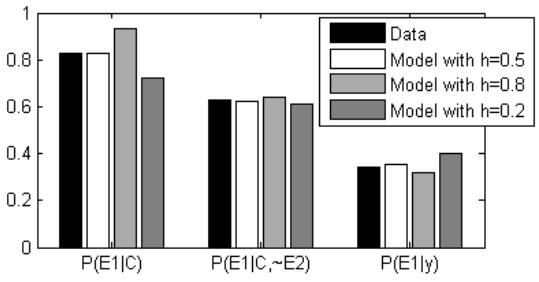


Figure 3: Data from Walsh and Sloman (2004), averaged over experiments, along with model predictions. Predictions are robust across alternative parameter settings.

After generating causal structures with branch points at various lengths, we used $h = 0.5$ and $a = 0.3$ to generate samples of the co-occurrence of the three events, by generating a set of replacements and propagating causation through the graph. Figure 3 shows the proportion of times that events occurred together, along with human probability judgments. We continued generating samples until we had at least 10000 samples for each entry. Predictions were resistant to changes in parameters; Figure 3 also shows other settings for $h$.

CERP can also easily make predictions about $P(E|\neg C)$, which Walsh and Sloman did not directly ask participants. However, they did ask participants a related question: the probability of an effect given a disabler that inhibited both effects (0.34). We sampled this by generating an active common inhibitor at a randomly chosen point $y \sim U[0,1]$, then choosing the branch point $x$ from $U[y,1]$, because it would be incoherent for the branch point to occur before the common disabler. This gave us a $P(E|y)$ of 0.35. This value is lower than $P(E|\neg C)$ because there is less of the path remaining on which a generative cause could fire.

Overall, the model explains the data well. Because there were few data points (three) in comparison to the number of parameters in the model (two), we will look at more experiments using the same parameters that best fit these data.

[2] This means: "$x$ was sampled from a uniform distribution between 0 and 1." We will use similar notation throughout the paper.

[3] We verified this by running the full generative model with a variety of parameter settings – several produced the same results as in Figure 3.

## Rehder and Burnett (2005)

Another dataset is provided by Rehder and Burnett (2005), who found a nonindependence effect in the domain of feature inference. They told adult participants that one novel feature of a category ($C$), caused three other novel features ($E_{1,2,3}$). They then asked participants to judge the probability of one feature in a member of the category, given some value of $C$, and some number of other collateral effects. In Experiment 1, participants were given a cover story involving natural kinds. In Experiment 2, participants were not given a cover story at all – they were told that abstract features caused each other. Across multiple experiments, participants predicted that the values of collateral effects would be correlated, showing a nonindependence effect even with no cover story. The results of their Experiments 1 and 2 are shown in Figure 4.

We modeled this much like Walsh & Sloman, 2004, except that there were two branch points and thus five edges. Again, we ensured that all paths from $C$ to $E_n$ had the same length, because participants were told that all effects had the same probability. All parameters were the same as in fitting Walsh and Sloman: $h = 0.5$, and $a = 0.3$. Because there was always one node with a longer branch than the others, we also randomly permuted the role of each node.

Results are shown in Figure 4. The model provides a good fit to the data, especially when the cause is present. In the absence of the cause, the model predicts slightly higher probability judgments than participants' responses. This is probably due to the effect of categorization: Other data show that participants in this paradigm were significantly less likely to believe that a given instance was actually a member of the category when $C$ was not present (Rehder & Burnett, 2005). This well-known "causal status effect" (Ahn, Kim, Lassaline, & Dennis, 2000) probably lowered their judgments of the other characteristic features of the category. Put another way, we assumed that feature $C$ was uncaused, but participants may have assumed that all features had a hidden common cause that was present only in category members. CERP can model the effect of such an additional hidden common cause, but that was not our goal in the present investigation. We model an experiment below that replicated Rehder and Burnett's findings outside the domain of categorization, where we do not find this problem.

## Mayrhofer, Hagmayer, and Waldmann (2008)

One strength of CERP is that it predicts how descriptions of the causal mechanism should affect the degree of nonindependence observed. Mayrhofer, Hagmayer, and Waldmann (2008) did just this. They told adult participants about four telepathic aliens; we will call them $C$, $E_1$, $E_2$ and $E_3$. The "cause" alien sometimes causes the "effect" aliens to think of food when he thinks of food. In the *transmit* condition, the instructions said that $C$ sent his thoughts to each $E_n$, but sometimes $C$ had difficulty concentrating. In the *receive* condition, instructions said that each $E_n$ read the thoughts of $C$, but each effect alien sometimes had difficulty concentrating.

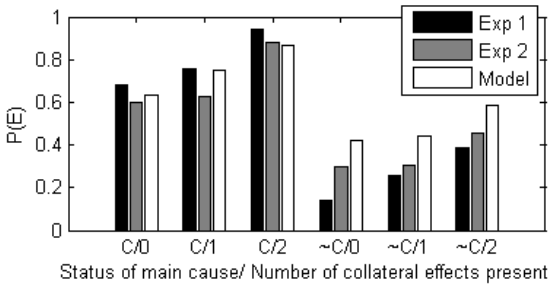


Figure 4: Data from Rehder and Burnett (2005) along with model predictions. Lower judgments in the absence of the cause feature are probably due to categorization, which is external to our model.

Like Rehder and Burnett (2005), they asked participants to judge $P(E)$ given different numbers of collateral effects. Data show a much stronger nonindependence effect in the transmit than in the receive condition (See Figure 5).

As before, we used a process which was equivalent to generating a large sample of graphs, and keeping only those consistent with the cover story. The cover story describes three similar effects, so we began by generating all three from the same branch point $x \sim U[0,1]$. We generated an inhibitor, explicitly mentioned in the cover story ("failure concentrating") at point $y \sim U[0,1]$, assigning it probability $a = 0.3$ of firing. In the receive condition, we kept only those samples in which $y > x$, since the inhibitor was described as applying to each alien individually. We generated one instance of the inhibitor on each branch. In the transmit condition, we kept only samples in which $y < x$, since only one inhibitor was described. We know of no other way to generate graphs consistent with both CERP and the cover story. Otherwise, we sampled as before, using $h = 0.5$ and $a = 0.3$.

As shown in in Figure 5, the model provides a good fit to the data. One exception is the point in the transmit condition in which two collateral effects occur, but the cause does not (the last entry in the "transmit" graph): The model predicted a medium probability judgment, while participants gave a low judgment. This may be due to a random anomaly in human responses, because the data are hard to explain under any account: As collateral effects were added, participants lowered, rather than raised, their probability judgment. This is not replicated in any other condition or experiment.

Mayrhofer et al. fit the qualitative difference between the conditions by adjusting a quantitative parameter: strength of inhibitory noise, which was strong in the transmit condition and weak in the receive condition. As they show, this parameter can be used to fit a wide range of data. Our model used a qualitative structural change instead, while the quantitative parameters have relatively little effect on the predictions, and remained constant between conditions and experiments. The model captures how changes to the mechanism description change the source and structure of the noise.

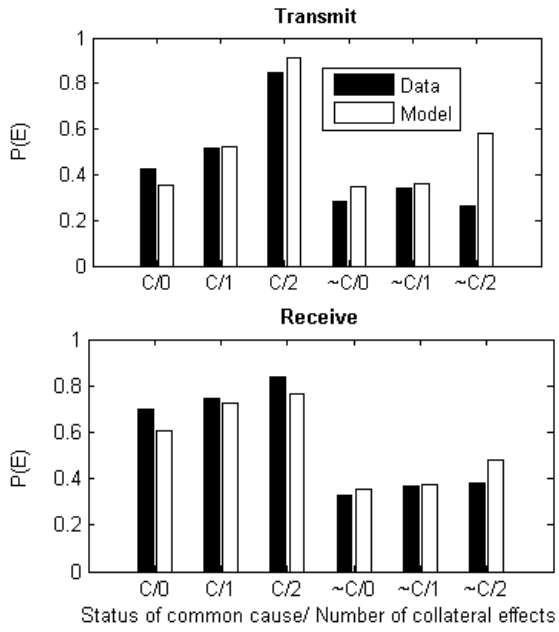


Figure 5: Data from Mayrhofer et al. (2008) along with model predictions. No parameters were varied between conditions –only the constraints given by the different cover stories.

## Conclusion and Further Work

We use CERP to fit three independently collected data sets on nonindependence, using the same parameters between experiments, and even between conditions within experiments. Over all three data sets, we fit 21 data points using 2 free parameters, with a correlation of greater than 0.99.[4] The power of CERP seems to come not from its use of free parameters, but from the fact that structural aspects may mirror some important aspect of the way that human beings represent causation. Further work will focus on exploring these aspects more closely. For instance, we can generalize our explanation for Mayrhofer et al., 2008's data to make a novel predictions: Early inhibitors in a causal stream should create more nonindependence than late inhibitors. We call this a *stream location effect*. We have recently tested this on preschoolers, with positive results (Buchanan & Sobel, 2010).

Our main intent with CERP is to test predictions that go well beyond nonindependence effects. For instance, its commitment to a form of determinism (namely, that apparent randomness always comes from hidden causes) has implications for how we reason about data that varies over time. Imagine your car fails to start one morning. Is it more likely to start tomorrow morning, or on a morning one year from now? If the relation were truly random, there should be no difference in judgment between these two times. If we introduce time into CERP, it should be able to rationally justify and fit our intuition that the car is more likely to start a year from now, than it is tomorrow. This is because variability arises from hidden causes that have persistence in time and space.

Finally, because it can generate any functional relation, CERP represents one way of defining a prior distribution over logical graphs. This may be useful to researchers (i.e. Lucas & Griffiths, 2010) who are interested in how people learn about the functional form of causal relations. An interesting question that arises from this research program is whether something like CERP could itself be learned – for instance, children might start with more general causal expectations, and come to realize that the world follows some or all of the commitments of CERP, such as determinism, and the stream-like character of causation.

[4] In data sets with multiple experiments, we correlated the model's predictions with the average over the experiments.

## Acknowledgments

Supported by NSF (DLS-0518161 to DMS). Thanks to Noah Goodman for discussions. Thanks also to Ralf Mayrhofer and Bob Rehder for generously sharing details of their data.

## References

Ahn, W., Kim, N., Lassaline, M., & Dennis, M. (2000). Causal status as a determinant of feature centrality. *Cognitive Psychology*, *41*(4), 361–416.

Buchanan, D., & Sobel, D. (2010). Causal stream location effects in preschoolers. In *Proceedings of the 32nd annual conference of the cognitive science society.*

Cheng, P. (1997). From covariation to causation: A causal power theory. *Psychological Review*, *104*(2), 367–405.

Keil, F. (2003). Folkscience: Coarse interpretations of a complex reality. *Trends in Cognitive Sciences*, *7*(8), 368.

Lucas, C., & Griffiths, T. (2010). Learning the form of causal relationships using hierarchical Bayesian models. *Cognitive Science*, *34*(1).

Mayrhofer, R., Hagmayer, Y., & Waldmann, M. (2008). Violations of screening off: A Bayesian error attribution model of causal reasoning. *Unpublished presentation at Mathematical Psychology 2008*.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press.

Rehder, B., & Burnett, R. (2005). Feature inference and the causal structure of categories. *Cognitive Psychology*, *50*(3), 264–314.

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

Shultz, T. (1982). Rules of causal attribution. *Monographs of the society for research in child development*, 1–51.

Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search*. The MIT Press.

Walsh, C., & Sloman, S. (2004). Revising causal beliefs. In *Proceedings of the 26th annual conference of the cognitive science society.*

Yuille, A., & Lu, H. (2008). The noisy-logical distribution and its application to causal inference. *Advances in neural information processing systems*, *20*, 1673–1680.