

# Children posit hidden causes to explain causal variability

David W. Buchanan (david\_buchanan@brown.edu) and David M. Sobel (dave\_sobel@brown.edu)

Department of Cognitive, Linguistic, and Psychological Sciences  
Box 1821, Brown University, Providence, RI, 02912

## Abstract

Most models of causal reasoning estimate the strength of a causal relation using a function of the proportion of successes and failures: the number of trials on which the cause produced the effect, divided by the total number of trials. Alternatively, people may represent failures as due to a hidden inhibitor that has a specific location and extent in time. We model these possibilities, and empirically test a case on which the two models make opposite predictions. We find that children as young as four years old generate responses inconsistent with proportional models, but consistent with an inhibitor-based model. Incorporating a recency component does not help proportional models fit the data.

**Keywords:** causal reasoning, cognitive development, models of causal reasoning, hidden Markov models.

## Introduction

Causal relations do not always hold. For instance, turning your key starts your car, but most of us have had an experience with a car that did not start when we turned the key in the ignition. Some of us own reliable cars, where this happens infrequently, whereas others of us own cars where this happens often. How do people represent such differences in variability?

One approach is to see variable causal relations as essentially probabilistic: causal relations sometimes randomly fail. In this case, it is natural to estimate the strength of the relation using some function of the proportion of successes to total trials. For instance, if your car started 6 days out of 10 in the past, you might estimate that the probability of it starting the next day, was 0.6. Most models from the causal reasoning literature either directly use, or converge to, such a proportion or a function of such a proportion. These include most associative (e.g., Rescorla & Wagner, 1972), causal power (e.g., Cheng, 1997) and neural network (e.g., McClelland & Thompson, 2007) models of causal reasoning. Not all of these models are deeply committed to using simple proportions, but most currently do. We will refer to such models as *proportional* models.

There is an alternative to the proportional approach: it could be that variability is the result not of intrinsic randomness, but rather the result of changing hidden inhibitors that tend to persist in time. For instance, it is possible that cold weather prevented the engine from turning over, or your battery needed to be replaced, or some other unknown problem existed. Under such an *inhibitor-based* model, judging the probability of success involves judging whether the causes responsible for past failures, might be active on the next trial. For instance, if your car usually starts six days out of ten, then you can reason that whatever hidden causes are responsible for the four failures out of ten, are 40 percent likely to occur

tomorrow. As this example illustrates, there are many cases in which proportional and inhibitor-based approaches make the same predictions. Given that inhibitor-based approaches introduce new complexity, it is understandable that researchers have focused on proportional models.

There are, however, some cases in which the two classes of models make different predictions. One of these cases occurs when we must reason about a novel intervention. For instance, imagine you own a new car that has never had a problem starting. On Monday, your friend, an amateur mechanic, opens the hood and makes some “improvements.” On Tuesday, your car fails to start. Without further repairs, it seems unlikely your car would start on Wednesday. On the other hand, imagine if the car was old junker, that normally only started half the time – it had failed on Thursday and Saturday, but started on Friday and Sunday. As above, your mechanic friend fiddles under the hood on Monday, and on Tuesday it fails to start. In this case, it seems less likely that the car will fail again on Wednesday. The fact that the car sometimes failed before, then recovered, makes you more confident that it will recover again.

Some experimental data suggests that children represent hidden inhibitors, when shown a novel, variable causal relation. Schulz and Sommerville (2006) showed 4.5 year-old children a machine that lit up and played music when a button was pressed. They also showed children a ring that when placed on the machine, appeared to prevent the machine from working. In the *probabilistic* condition, children then saw a period in which the machine activated on only a few trials, in the following pattern: [0 1 0 0 0 1 0]<sup>1</sup> In the *deterministic* condition, children saw eight successes: [1 1 1 1 1 1 1]. At test, children in both conditions were presented with a choice: they could disable the machine using the ring (which they had seen work as a disabler before), or they could disable the machine using a candidate cause (a small flashlight) that had been hidden during the rest of the procedure. Children were significantly more likely to choose the flashlight in the probabilistic than in the deterministic condition. This result suggests that the probabilistic activation implied, to the children, the existence of a hidden inhibitor. A proportional model cannot explain this effect, since the calculation of causal strength in both conditions should be qualitatively the same.

Schulz and Sommerville (2006)’s data provide support for a model that explicitly incorporates hidden inhibitors. To our knowledge, no experiment exists that directly compares the predictions of inhibitor-based, and proportional approaches.

<sup>1</sup>This notation means: one failure, one success, then four failures, followed by one success, followed by one failure. This notation will be used throughout the paper.

We sought to create such a scenario. Specifically, inhibitor-based models predict that in some cases (such as the car example above) adding failures to a string of instances, should make it judged more likely that a success will occur subsequently. This is because the consistency of successes can allow us to infer the likely consistency of failures. Proportional models, as we will see below, cannot match these predictions. We will begin by describing the methods and results of the experiment, before describing the models and model predictions. We will then discuss the implications of these results for models of causal reasoning.

## Methods

### Participants

Participants included 32 three-year-olds ( $M = 41.50$  months, Range: 36-47 months, 12 girls), and 32 four-year olds ( $M = 53.3$  months, Range: 48-59 months, 18 girls). Four additional children (all three-year-olds, two in each condition) were tested, but excluded because they failed controls (described below). Half the included children in each age group were randomly assigned to the *deterministic* condition, while the other half were assigned to the *probabilistic* condition. About half of the children were tested in a laboratory at Brown University, while the remaining children were tested in a quiet room at the Providence Children's Museum.

### Materials

Materials included a set of six commercially available closet lights. Each resembled a large button that illuminated when pressed. (See Figure 1.) One of the lights, which we will call the "effect light", measured 14 cm in diameter. It was modified such that it could be activated surreptitiously using a remote. Inside of the battery compartment, we installed a dummy switch that was not connected to the wiring of the light. A second light (the "decoy light") had the same appearance, but never activated when pressed. Four more lights (the "cause lights") were smaller, each measuring 10 cm in diameter, and were each painted a different color. These lights were modified such that they always illuminated when the button was actively depressed, and stopped illuminating when the button was released. The lights were mounted together on a piece of cardboard, apparently connected with black wires, as shown in Figure 1.

This setup allowed the experimenter to create the illusion, using the remote, that the cause lights actually caused the effect light to illuminate. The experimenter, a trained magician, practiced this effect until a convincing causal impression was achieved. In addition, a set of colored pipe cleaners that could be wrapped around the effect light, allowed the experimenter to change the apparent identity of the effect light, without needing to create multiple remote-controlled lights. No children noticed either of these deceptions during the experiment.

### Procedure

**Training Phase** The experimenter first brought out the effect light, and demonstrated that pressing on the effect light

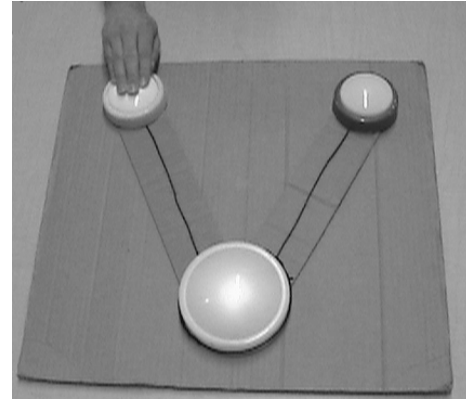


Figure 1: The lights used in the present experiment, as they were shown to children in the test phase.

caused it to activate. The children were allowed to press the light as well, which illuminated when they pressed it. (This was achieved using the remote.) The experimenter then held his hand over the light, and asked: "I'm going to press on this light again. Will it light up?" If children answered correctly, saying "yes," the experimenter moved to the next part of the experiment. If they answered incorrectly, gave a nonverbal response, or did not answer, children were given corrective feedback, and the experimenter pressed the light again, and asked the question again. If children needed more than three instances of corrective feedback, their data were excluded from the experiment for failing the controls. The experimenter repeated the training phase with the decoy light, which did not illuminate. In this case, the correct answer was "no." Note that at this stage, all children who passed controls had correctly answered "yes" to one question, and "no" to another. Thus, children using the strategy of answering "yes" to every question, were excluded.

**Test Phase: Deterministic Condition** The experimenter brought out three apparently new lights, and arranged them as in Figure 1. He said: "Sometimes, when you push on the little lights, they make the big light go. Let's see what they do." The experimenter first pressed three times on the light to the child's right, which apparently caused the effect light to illuminate each time. He then pressed three times on the left light, with the same result. Finally, he pressed on the right light three more times, again causing the effect light to illuminate each time. The experimenter then picked up the effect light, and said "I'm just going to do something to the light." He held the light vertically, such that the underside was hidden to the child, opened the back cover, and flipped a switch inside of the battery compartment with an audible click. (The switch had no real electronic effect or purpose.) He then replaced the light on the cardboard, and pressed on the right light. The right light illuminated, but the effect light did not. The experimenter then asked, pointing to the right light: "I'm going to push on this one again. What do you think: If I push on this one again, will it make the big light go?" He then

## Experimental Results

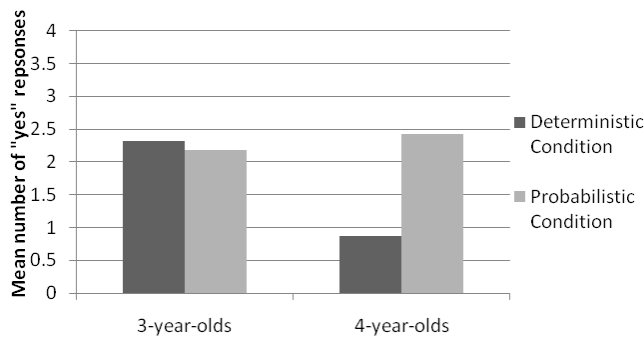


Figure 2: Experimental results. Four-year-olds were more likely to say “yes” in the probabilistic than in the deterministic condition.

repeated the question, referring to the left light. Children’s answers were recorded. The experimenter then repeated the whole test phase with an apparently new set of three lights, starting with the light on the child’s left. This yielded a total of four yes/no questions from each child. The color of the lights were randomly counterbalanced, as was the starting location of the light referred to here for simplicity as the “right” light. That is, about half the children saw the experimenter start with the light on their left, and about half the children saw the experimenter start with the light on their right.

**Test Phase: Probabilistic Condition** Children in the probabilistic condition were given the same procedure, except that the effect light sometimes failed to illuminate when the cause lights were pressed, even before the experimenter flipped the switch. (The cause lights always illuminated when pressed.) We can see the deterministic condition as a series of nine successes, followed by an intervention, followed by a failure, followed by a trial on which a prediction was requested. We will represent this as: [1 1 1 1 1 1 1 1 — 0 ?]. The probabilistic condition always followed the pattern [1 0 1 1 1 0 0 1 1 — 0 ?]. That is, the right light succeeded, failed, then succeeded, the left light succeeded, then succeeded, then failed, and finally the right light failed, then succeeded, then succeeded. This sequence was constructed to appear random, containing an appropriate number of alternations, but it also terminated with two successes before the intervention. Note also that the total proportion of successes was the same for each light (2/3).

## Experimental Results

We began by assigning each child a score, which was the total number of times that the child answered “yes” to the test questions (range: 0 to 4). Means are shown in Figure 2. Preliminary analyses revealed no effects of gender, or counterbalanced factors such as the side of the first demonstrated light. We subjected the data to a two-way (age  $\times$  condition) ANOVA. This revealed a main effect of condition,

( $F(1, 60) = 5.71, p = 0.020$ ), no main effect of age group ( $F(1, 60) = 2.83, p = 0.098$ ), and a significant interaction between age group and condition ( $F(1, 60) = 4.148, p = 0.046$ ). Because of the interaction, we investigated simple effects of condition within each age group. There was a significant effect of condition for 4-year-olds, ( $t(30) = 3.30, p = 0.002$ ), but not for 3-year-olds ( $t(30) = 0.239, p = 0.81$ ). This difference occurred because 4-year-olds were more likely to say “yes” in the probabilistic, than in the deterministic condition.

Subsequent analyses verified these patterns by looking at individual questions. Children were asked four questions: about the light they had seen fail after the intervention on the first trial (“demo-1”), the light whose efficacy they had not seen after the intervention (“other-1”) and the corresponding lights on the second trial (“demo-2” and “other-2”). Within each age group, we subjected the pattern of yes/no responding on each question between conditions, to a two-sided Fisher’s exact test. Three-year-olds showed no significantly different pattern of responding, on any of the four test questions (all  $p$ -values greater than 0.05). Four-year-olds, however, showed significantly more “yes” answers in the probabilistic than the deterministic condition, on three out of the four questions: demo-1 ( $p = 0.02$ ), demo-2 ( $p = 0.015$ ), and other-2 ( $p = 0.011$ ). They did not show a significant difference on other-1 ( $p = 1.00$ ). This may be because some children felt the need to balance their initial “no” response, with a “yes” response. Importantly, this did not seem to happen on the second trial (when presumably it became obvious that most answers were “no”). Most importantly, note that 4-year-old children showed a significant difference on the first question asked.

Chance analyses were in accord with these results. Chance was defined as a 50 percent probability of saying “yes”. Binomial tests on the pattern of responding on each question, within each age group and condition, did not reveal any questions on which 3-year-olds responded significantly differently from chance. The same was true of 4-year-olds in the probabilistic condition. In the deterministic condition 4-year-olds were significantly more likely to answer “no” than chance would predict, on demo-1 ( $p = 0.001$ ), demo-2 ( $p = 0.001$ ), and other-2 ( $p = 0.021$ ), but not on other-1 ( $p = 0.21$ ).

Overall, 4-year-olds showed a clear tendency to respond “yes” more often in the probabilistic condition, than in the deterministic condition. Three-year-olds did not show this tendency – they were not significantly different from chance in either condition.

## Model Description and Results

In this section, we will describe the predictions made by the two classes of models. The prediction we will extract from each class of models is on which condition, if any, children should be most likely to say “yes” to the test question. To preview, each class of model will make internally unanimous predictions, but the predictions will be different between proportional and inhibitor-based models.

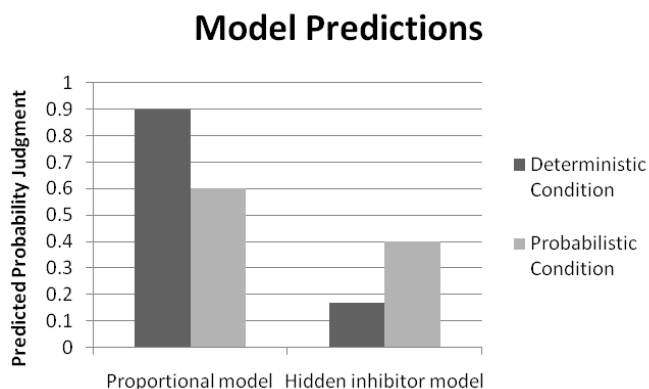


Figure 3: Model predictions. Only the hidden inhibitor model is consistent with the 4-year-old data.

Some assumptions will be common to both models. The first assumption is that the probability of the effect in the absence of the cause is zero. This is natural given that children never see the effect light illuminate in the absence of pressing the cause lights. Second, both models will assume for simplicity that children should answer “yes” to the test question if the probability of the effect occurring is greater than some threshold. Preschoolers like to say “yes,” (Heather Fritzley & Lee, 2003), so it is plausible that this threshold is lower than 50 per cent. We will assume that higher probabilities generally correspond to more frequent “yes” responses. Because other factors (such as guessing) may go into children’s responses, we will not treat chance-like responding as inconsistent with any model. For instance, adding noise to any model can produce chance-like responding.

We begin with the class of proportional models, by first considering the simplest possible proportional model: the probability of success is the proportion of successes so far. In the deterministic condition, children see nine successes, and one failure, meaning that the probability of success on the 11th trial is 0.9. In the probabilistic condition, children see six successes, and four failures, meaning that the probability of success is 0.6. The proportion is lower in the probabilistic condition. Therefore, a basic proportional model is committed to the prediction that children should be more likely to say “yes” in the deterministic than probabilistic condition. Predictions are shown in Figure 3.

These predictions are the opposite of the experimental results from 4-year-olds. Can we find a proportional model that fits the 4-year-old data? One approach would be to throw out all data from before the intervention, and learn only from the one post-intervention trial. This could be reasonable if children believe that the intervention changed the system. Because the post-intervention trial was identical between conditions, this commits the new proportional model to the prediction that the number of “yes” responses between conditions should be the same. While it is possible that 3-year-olds are reasoning in accord with this model, they could also be

guessing. The 4-year-old data were inconsistent with these predictions.

Another approach might be to give recent trials more weight than trials earlier in the sequence. This could be reasonable both psychologically (fading memory) and rationally (recent trials are more relevant). We can use a function to assign a different weight to each trial, presumably with more weight assigned to recent trials.<sup>2</sup> The predicted proportion is the total weight of the successful trials, divided by the total weight of all the trials. There were some trials on which there was a failure in the probabilistic condition, but not in the deterministic condition. Therefore, in order to fit the data, the recency function would have to discount these failures through the weight function. It would also have to assign a high weight to at least one trial in which there was a success in the probabilistic condition, but a failure in the deterministic condition. However, no such trial exists: every failure in the deterministic condition is also a failure in the probabilistic condition. Put another way, the experimental effect shows that adding failures to a sequence, while holding successes constant, can increase probability judgments. Proportional models cannot capture this effect, because adding failures should decrease judgments, if they have any effect.

We propose an inhibitor-based model that makes different predictions. It takes a generative approach, describing a set of rules for randomly creating a sequence of successes and failures. The probability of a set of sequences given a condition, is equal to the number of times we expect to generate that set of sequences, divided by the number of times we expect to generate all the sequences that meet the condition.

The inhibitor-based model accounts for variability in the causal relation, by positing the existence of a hidden inhibitor whose state changes over time. We begin by considering only one inhibitor for simplicity; we will consider multiple inhibitors below. The inhibitor can be either ON or OFF; when it is ON the causal relation does not hold. The basic model has two parameters:  $0 < a < 1$ , which describes the overall proportion of time that the inhibitor is active and  $v > 0$ , which describes the volatility of the system: how likely the inhibitor is to change state. The model starts setting the inhibitor to ON on the first trial with probability  $a$ . At each time step, change the inhibitor from ON to OFF with probability  $(1 - a)(1 - e^{-v})$ , or from OFF to ON with probability  $a(1 - e^{-v})$ . This model can generate a sequence of successes and failures of arbitrary length: successes occur if and only if the inhibitor is OFF.

Some properties of the model are worth noting. First, note that in the long run, the steady state of the Markov chain defined by the two transition probabilities, is  $[a, (1 - a)]$ . This achieves the desired property that the inhibitor is ON a proportion of time defined by  $a$ . Second, note that for any finite value of  $v$ , there exists a statistical dependence between the value of the inhibitor at each time step: it is more likely to be

<sup>2</sup>Note that the argument in this paragraph applies to all possible weight functions, not just ones that decrease monotonically.

on at time  $t + 1$  if it is on at time  $t$ . As  $\nu$  approaches infinity, the correlation between time steps approaches zero. Note that  $\nu$  is not allowed to equal zero because the inhibitor would never change state. However, as  $\nu$  is reduced toward zero, the dependence between time steps becomes stronger. The model is derived from considering a continuous underlying Poisson process in which change events arrive at rate  $\nu$ . At each event, we re-sample the value of the inhibitor from  $a$ . Because of this underlying process, we can discretize time differently by using different values of  $\nu$ . For instance, to model time steps that are twice as long, we can double  $\nu$ .

To generate model predictions, we ran a sampler that followed the algorithm described above, with  $a$  set to 0.5, and  $\nu$  randomly sampled each time from an exponential distribution with mean 1. At each step, the sampler generated a sequence of 11 successes and failures. We kept only those samples in which the first ten trials exactly matched the pattern of activation for the given condition. The estimated probability judgment is the proportion of these samples in which success occurred on the 11th trial. For instance, in the probabilistic condition, the predicted probability judgment is the number of times [1 0 1 1 1 0 0 1 1 0 1] occurred, divided by the number of times that [1 0 1 1 1 0 0 1 1 0 1] and [1 0 1 1 1 0 0 1 1 0 0] occurred combined. The sampler ran until it generated at least 10000 samples that matched the pattern of activation.

Results are shown in Figure 3. Note that the model predicts a higher probability judgment in the probabilistic condition than in the deterministic condition. Unlike the proportional model, the inhibitor-based model is consistent with the pattern of data observed in 4-year-olds.

We can interpret these results as follows: cases such as the probabilistic condition, in which a highly variable sequence is generated, tend to have high volatility. This means that a failure is moderately likely given a failure. However, cases like the deterministic condition, where a highly consistent sequence is generated, tend to have low volatility: a failure is highly likely given a failure, since a success was highly likely given a success. In other words, children should infer the consistency of failures from the consistency of successes. This should be true even in cases (such as the deterministic condition) when they have never seen a sequence of failures.

The model just presented is only the simplest version of the hidden inhibitor model. We replicated the qualitative predictions of this model in several more sophisticated versions of the hidden inhibitor model. One included a model in which  $a$  was re-sampled from a uniform distribution at each sample. In another case the sampler generated multiple inhibitors (a number of inhibitors chosen randomly from  $Poisson(1)$  at each sample) and success occurred only if every inhibitor was off on a given trial. We also constructed a model that explicitly incorporated the intervention: one of the inhibitors was randomly changed between the 9th and 10th trial (corresponding to the time we made the intervention in the experiment). In each case, the prediction held: children should be more likely to say “yes” in the probabilistic than in the

deterministic condition.

Readers familiar with the literature on randomness perception (e.g. Oskarsson, Van Boven, McClelland, & Hastie, 2009) will notice parallels between the hidden inhibitor model presented here, and models of randomness perception. There is clearly overlap between the two fields: participants in randomness perception experiments often perceive random sequences as causal. While hidden Markov models are used in randomness perception, we do not intend this inhibitor-based model as a model of random sequence generation. The inhibitor-based model is intended specifically as a model of causal, not all random, events.

Overall, both classes of models are robustly and unambiguously committed to a difference between the two conditions, but in opposite directions. The class of proportional models is committed to judgments that are at least as high (and in most cases, higher) in the deterministic than in the probabilistic condition. Conversely, the class of hidden inhibitor models is committed to judgments that are higher in the probabilistic condition.

## Discussion

Experimental results showed that 4-year-olds were more likely to predict a failure given a consistent series of successes followed by a failure, than in an otherwise identical condition in which some of the successes were replaced by failures. Three-year-olds were not significantly different from chance in either condition. Data from 3-year-olds were inconclusive, because they could be consistent with weak and/or noisy versions of either model. Data from 4-year olds contradict the predictions of proportional models, but are consistent with hidden inhibitor models.

One possible objection to these conclusions is that we have considered multiple causes of the same effect. That is, the left and right lights are treated as the same cause for modeling purposes. To be sure that this did not bias our results and predictions, we repeated the analysis using just the data from the right light. We considered only the six trials on which the right light was the cause, and ignored the three trials on which the left light was used. Model predictions were qualitatively the same. Note that the experimental results are the same as well – 4-year-old’s overall pattern of responding was consistent with their responses to the first test question, which referred to the right light.

An alternative explanation for these data uses a simpler model than the one presented here: Children expect causal relations to always hold, unless a human being intervenes on an inhibitor. Children may have guessed in the probabilistic condition, because they were baffled when the relation failed without apparent intervention. While this explanation is simpler, and explains the data, we believe that it is at odds with other considerations about children’s causal reasoning. For instance, children are exposed to toys that fail sometimes without human intervention: batteries run out, weather conditions affect television reception, etc. For this reason, we

endorse a more sophisticated model that involves volatility, like the one presented here.

The developmental difference was unexpected, but interesting. Why do 3-year-olds respond differently from 4-year-olds? Because 3-year-olds did not show significantly more “yes” responses than chance would predict, we cannot use a yes-bias to explain the data. Furthermore, three-year-olds were able to readily overcome any yes-bias when this paradigm was used in other experiments. For instance, Buchanan and Sobel (2010) observed a significant number of “no” answers from 3-year-olds using the same set of three lights, in the same arrangement, in a different experiment.

One possibility is that differences in memory and processing capacity are responsible for the developmental difference. For instance, 3-year-olds may have lost track of the pattern of successes and failures, and simply guessed. Further work could test this explanation, by presenting children with a visual representation of the pattern of successes and failures, or by using a shorter sequence. At present, there is insufficient data to rule out this explanation.

Finally, it could be that 4-year-olds are more likely to represent hidden inhibitors than 3-year-olds. This is broadly consistent with a number of studies that suggest developmental differences exist between 3- and 4-year-olds in their understanding of causal mechanisms (Buchanan & Sobel, in press; Bullock, Gelman, & Baillargeon, 1982) and the relation of causal properties to hidden internal properties (Sobel, Yoachim, Gopnik, Meltzoff, & Blumenthal, 2007). As far as we know, the present experiment is the first to show a developmental difference in positing the presence of hidden inhibitors. Further experiments are needed before we can speculate about the exact cause and nature of the developmental difference.

Even outside of developmental considerations, these data and model present a challenge for theories of how human beings represent causal relations. Partly because of their simplicity, proportional approaches to causal strength have persisted through several generations of models. They are used in most associative, causal power, and some Bayesian models of causal representation. The present data suggest that even young children may use a more complex representation of causality, at least in some cases. If these results can be replicated and extended, then they may form the basis of improved models. Elsewhere (Buchanan, Tenenbaum, & Sobel, 2010) we present one way of incorporating hidden inhibitors into a Bayesian framework. We expect that inhibitor-based models could be incorporated into multiple modeling paradigms.

If inhibitor-based models are necessary when modeling causal relations, this may have theoretical implications for causal reasoning in general. For instance, people may be motivated to represent inhibitors because of an implicit determinism (Schulz & Sommerville, 2006): causal relations are never intrinsically random, so variability must be the result of hidden causes. Furthermore, if people represent hidden inhibitors, they may represent causal mechanisms – the events

that happen between cause and effect – in more detail than some theories currently acknowledge (for a discussion, see Ahn, Kalish, Medin, & Gelman, 1995). Further work will explore theoretical connections to these questions.

## Acknowledgments

This work was supported by NSF (DLS-0518161 to DMS). We would like to thank all of the parents and children who participated, and the Providence Children’s Museum for allowing us to recruit and test on-site. We also wish to thank Deanna Marcis and Katie Green for help with data collection and analysis.

## References

- Ahn, W., Kalish, C., Medin, D., & Gelman, S. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition*, 54(3), 299–352.
- Buchanan, D., & Sobel, D. (2010). Stream location effects in preschoolers. In *Proceedings of the 32nd annual conference of the cognitive science society*.
- Buchanan, D., & Sobel, D. (in press). Mechanism-based reasoning in young children. *Child Development*.
- 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*.
- Bullock, M., Gelman, R., & Baillargeon, R. (1982). The development of causal reasoning. In W. Friedman (Ed.), *The developmental psychology of time* (pp. 209–254). New York: Academic Press.
- Cheng, P. (1997). From covariation to causation: A causal power theory. *Psychological Review*, 104(2), 367–405.
- Heather Fritzley, V., & Lee, K. (2003). Do young children always say yes to yes–no questions? A metadevelopmental study of the affirmation bias. *Child Development*, 74(5), 1297–1313.
- McClelland, J., & Thompson, R. (2007). Using domain-general principles to explain children’s causal reasoning abilities. *Developmental Science*, 10(3), 333–356.
- Oskarsson, A., Van Boven, L., McClelland, G., & Hastie, R. (2009). What’s next? Judging sequences of binary events. *Psychological Bulletin*, 135(2), 262.
- Rescorla, R., & Wagner, A. (1972). Variations in the Effectiveness of Reinforcement and Nonreinforcement. In *Classical conditioning ii: Current research and theory*. New York: Appleton-Century-Crofts.
- Schulz, L., & Sommerville, J. (2006). God does not play dice: Causal determinism and preschoolers’ causal inferences. *Child development*, 77(2), 427–442.
- Sobel, D., Yoachim, C., Gopnik, A., Meltzoff, A., & Blumenthal, E. (2007). The blicket within: Preschoolers’ inferences about insides and causes. *Journal of Cognition and Development*, 8(2), 159–182.