# Analytic Causal Knowledge for Constructing Useable Empirical Causal Knowledge: Two Experiments on Preschoolers

**Patricia W. Cheng (cheng@lifesci.ucla.edu)**
Department of Psychology, University of California, Los Angeles, CA 90095 USA

**Mimi Liljeholm (m.liljeholm@uci.edu)**
Department of Cognitive Sciences, University of California, Irvine, CA 92697 USA

**Catherine M. Sandhofer (sandhof@psych.ucla.edu)**
Department of Psychology, University of California, Los Angeles, CA 90095 USA

## Abstract

The present paper examines what domain-general causal knowledge reasoners need for at least some outcome-variable types to construct useable content-specific causal knowledge. In particular, it explains why it is essential to have analytic knowledge of causal-invariance integration functions: knowledge for predicting the expected outcome assuming that the empirical knowledge acquired regarding a causal relation holds across the learning context and an application context. The paper reports two studies that support the hypothesis that preschool children have such knowledge regarding binary causes and effects, enabling them to generalize across contexts rationally, favoring the causal-invariance hypothesis over alternative hypotheses, including interaction (e.g., linear) integration functions, heuristics, and biases.

**Keywords:** Causal induction; causal learning; causal invariance, rationality; cognitive development

## Introduction

How do we humans best represent the world so that we are able to achieve desired outcomes? A basic requirement is that the world knowledge we acquire be useable. Whenever we use our past knowledge to achieve a desired outcome (e.g., avoid a certain food to prevent a skin or intestinal reaction), we are inevitably generalizing from the learning context (e.g., items for meals at home preceding past allergic reactions) to a subsequent new context (lunch at work the next day, food during foreign travel). By *contexts* with respect to a cause in question, we mean occasions or settings where (known or unknown) enabling conditions and alternative causes of the target outcome may occur with different probabilities.

By adulthood, humans appear to make causal judgments that suggest they assume *causal invariance* – namely, that causes operate in an invariant manner across the learning and application contexts – as a default and as a criterion for revising causal knowledge (e.g., Cheng, 1997; Liljeholm & Cheng, 2007; Lu, Rojas, Beckers & Yuille, 2016; Lu, Yuille, Liljeholm, Cheng & Holyoak, 2008). If the concept of causal invariance is essential to the construction of useable causal knowledge, we would expect young children to use it just as adults do. Alternatively, if the concept operates as an acquired strategy or heuristic in causal-knowledge construction, young children would be less likely to use it, especially when its use requires mathematical skills that are far beyond the children's general level of mathematical capability.

A large literature on children's causal reasoning shows that children are able to reason causally from a young age (e.g., Gopnik, 2009; Gweon & Schulz, 2011; Legare, 2012; Rakison & Krogh, 2012). For example, like adults, children can learn deterministic conjunctive or disjunctive causal relationships and generalize the relationship that better fits the evidence to other variables (Lucas, Bridgers, Griffiths & Gopnik, 2014). However, there has been little work on the essentiality of the causal-invariance concept in the shaping of causal knowledge. In particular, it is not known whether children use that concept, rather than simple approximations or heuristics. Young children's use of a probabilistic form of causal invariance would provide especially strong support for its essentiality.

To see why knowledge of causal invariance is essential for constructing useable causal knowledge, consider situations in which a) there may be background causes present, b) these causes may vary from context to context, and c) the set of candidate causes under evaluation may not include one that generalizes well across contexts. Natural settings often hold these challenges. When we want to infer what cures an illness, for example, the illness must have some non-zero probability of occurring due to some background generative cause. The illness need not occur across all individuals, suggesting that background alternative preventive causes may be present. And the illness may be more or less prevalent in different contexts (e.g., countries). The fact that the "no confounding" condition is a standard principle in experimental design is an indication of the pervasive need for the influence of a target cause to be teased apart from that of background causes. A further challenge is that our initial parsing of events to isolate distinct candidate causes may not yield predictions that generalize to application contexts. Moreover, generalizability is a matter of degree (Woodward, 2000, 2003). We may encounter occasions on which a relation that we have assumed to be generalizable unexpectedly fails to hold (e.g., when on a trip up a tall mountain we find that eggs boiled the usual amount of time remain uncooked). The replicability crisis in medical research is a reminder of failures to generalize even in costly planned investigations, not to say in everyday inferences. The need to go beyond one's current set of candidate causes is ever present.

Given the goal of formulating useable causal knowledge, information about a failure to reach that goal – failure indicated by a notable deviation from the outcome expected assuming that the acquired knowledge generalizes when applied – would be useful for assessing whether to retain or revise that knowledge. Along with our colleagues, we have proposed that mathematical functions characterizing the sameness of influence of a cause across contexts – functions which differ depending on the form of the cause and effect variables (e.g., binary vs continuous) rather than their content (e.g., tobacco smoking causes lung cancer) – play a critical role in the construction of causal knowledge (e.g., Cheng, Liljeholm & Sandhofer, 2013; Cheng & Lu, in press). We term these *causal-invariance functions*. Whenever there are too many possible causal models to exhaustively evaluate, causal invariance is a helpful signal.

We have further noted that the vastness of the search space of possible causal representations renders the use of causal invariance not merely helpful but essential. A basic tenet of cognitive science -- that our perception and conception of reality are our representations -- implies that the search space of the representation of reality is infinite. In an infinite search space, an exhaustive evaluation of the possible causal models is not merely practically infeasible, but in principle impossible. Given the nature of the problem of causal knowledge construction, the need to go beyond one's current candidate causes becomes clear. Deviation from the outcome expected based on causal-invariance functions serves as an essential navigating device.

What if the need for revision is signaled instead by deviation from a *causal-interaction* (i.e., non-causal-invariance) criterion? In that case, that is, if candidate *c*'s influence on target effect *e* is expected to vary depending on the state of the background causes, there would be a deviation from expectation -- signaling a need to revise causal knowledge -- when the influence of *c* in fact generalizes across contexts. Conversely, no deviation from expectation would confirm that *c* interacts with background causes (its inferred influence therefore should not generalize across contexts). But no deviation from expectation means no signal to revise. Given an inverted signal to revise, in the infinite search space of possible representations of reality, the acquired causal knowledge is unlikely to hold when applied or to replicate when further tested.

If our thesis on the essentiality of the concept of causal invariance is correct, we would expect young children to use the concept, even when its use requires mathematical skills that are far beyond the children's general level of mathematical competence, and even though such usage contradicts an irrational but common practice in medical or business research. Our two studies on preschool-aged children tested their use of a causal-invariance versus a causal-interaction criterion.

## Analytic Knowledge of Causal Invariance

For all situations, every observed outcome is inherently the outcome due to the totality of its causes; the contributing causal relations are not differentiable by observation. When background causes are present, the unobservability of causation requires that causal learners adopt an assumption (either tacitly or explicitly) regarding how the total causal influence that results in the observed outcome is *decomposed* into the influences by the candidate and the background causes. The functions characterizing the decomposition are often called *integration functions*. Causal invariance functions are integration functions that specify the sameness of causal influence across contexts. Different integration functions yield different causal conclusions (e.g., see Lu et al, 2008). Our Study 1 presents a situation where multiple integration functions yield qualitatively different causal recommendations.

One might argue, however, "Why would a particular integration function have a special status? Which integration function is appropriate depends on the domain. Although causal-invariance functions explain the results from many experiments (e.g., see Lu et al., 2008), perhaps due to reasoners' prior knowledge of how some causes combine their influences in certain scenarios, other integration functions may be more appropriate for describing how causal influences combine in other domains." Even if causal-invariance functions are the default integration functions, the argument may go, "whenever these functions do not fit the data from a domain, they would be – and should be – given up in favor of a better-fitting integration function. Causal-invariance functions may be a convenience, but the key factor is how well an integration function explains causation in a domain. Adherence to particular integration functions regardless of domain would be irrational." This argument may appear to have empirical support: Adults and even children have been shown to be able to learn various causal integration functions and generalize their learning to novel variables presented in the experiments (e.g., Lucas et al., 2014; Melchers et al., 2004).

To explain the relation between our work and work on integration-function learning, we make two distinctions: 1) a distinction between *analytic* and *empirical* knowledge (cf. Hume's, 1739, "truths of reason" and "matters of fact") and 2) a part-whole distinction, between a "*whole*" cause (elemental or complex) and an *interactive component* within a whole cause. Whereas empirical knowledge is content-specific and justified by experience or data, analytic knowledge is content- and domain-general (i.e., formal) and is justified by reason, by what deductively follows based on the meaning of the concepts in question. Previous work has studied the generalization of acquired empirical (data-based) integration functions. In contrast, our work studies the role of a causal-invariance function as analytic knowledge, operating as a default and a revision criterion in causal-knowledge construction, with both roles motivated by the (tacit) goal of formulating useable causal knowledge.

The combination of biological factors that lead to "healthy forest growth" is a whole cause of that outcome; adequate nitrates in the forest soil is an interactive component in that complex whole cause. Arsonists and the lumber industry are two other whole causes that influence that outcome. Likewise, the gravitational force from a celestial body $y$ on a celestial $x$ is a whole cause of $x$'s motion; the masses of bodies $x$ and $y$ and the distance between them are interactive components within that whole cause. The gravitational forces from other celestial bodies on $x$ are other contributing whole causes of $x$'s motion, independently influencing that motion.

Note that within the same domain (e.g., gravitational force), an interaction function (i.e., Newton's law of universal gravitation) integrates the influences from specific component factors (e.g., the masses of the two celestial bodies in a pairwise gravitational force) and a causal-invariance function (vector addition) integrates the influences from multiple whole (presumably non-interacting) causes (e.g., the gravitational forces from multiple bodies on a target body simply sum up). To enable prediction, the aim of causal-knowledge construction is to formulate whole causes (elemental or complex) that are teased apart from, that do not interact with, other causes (e.g., whole causes in the background).

Because causal-invariance and causal-interaction functions exist within the same domain, empirical integration functions are content- or context-specific rather than domain-specific. Whether an acquired interaction-integration function generalizes to other candidate causes depends on the perceived similarity between the relevant causal mechanisms (e.g., Lucas & Griffiths, 2010, Expt. 5; Wheeler, Miller, and Beckers, 2008, Expt. 3) as well as on situational variables (e.g., Wheeler et al., 2008, Expts 1 & 2; the demand characteristics of an experiment). In contrast, causal-invariance functions (e.g., vector addition, the noisy-AND-NOT function in Eq. 2) are formal, specific to variable types (vectors & binary variables, respectively), but general across domains, contents, and contexts. As explained earlier, for the goal of constructing useable causal knowledge, only causal-invariance functions can serve as a default and a revision criterion for integrating the influence of ideally whole candidate causes with the influence of (potentially unknown) other causes.

## Causal-Invariance Functions for Binary Variables

The causal-invariance functions for two binary causes of a binary effect – a candidate cause of an outcome and the background causes as group – are as follows (e.g., Cheng, 1997; Pearl, 1988). There are different but logically consistent functions for potentially generative and potentially preventive candidate causes.

For a candidate cause $c$ that potentially generates effect $e$ and does so independently of alternative causes in the context, denoted $a$ as a group, the probability of observing $e$ is given by a "noisy-OR" integration function,

$$P(e=1|c;w_a,q_c) = q_c \cdot c + w_a - q_c \cdot c \cdot w_a \quad (1)$$

where $c \in \{0,1\}$ denotes the absence and the presence of candidate cause $c$, $e \in \{0,1\}$ denotes the absence and the presence of effect $e$, $q_c$ represents the generative power of the candidate cause $c$, and $w_a$ represents the probability that $e$ occurs due to all background causes, known and unknown. For a candidate cause $c$ that potentially prevents effect $e$, the probability of observing $e$ is given by a "noisy-AND-NOT" integration function:

$$P(e=1|c;w_a,p_c) = w_a(1-p_c \cdot c) \quad (2)$$

where $p_c$ is the preventive causal power of $c$. These "noisy-logical" integration functions (terminology due to Yuille & Lu, 2008), under the assumption that there is *no confounding* [i.e., when P($a$ = 1|$c$ = 1) = P($a$ = 1|$c$ = 0)], imply respectively equations for estimating $q_c$ and $p_c$. The equation for estimating preventive power $p_c$, for example, is:

$$p_c = \frac{P(e=1|c=0) - P(e=1|c=1)}{P(e=1|c=0)} \quad (3)$$

Our experiments test preschoolers' use of noisy-logical functions, the probabilistic version of disjunction, in their role as analytic knowledge of causal invariance for binary variables. Testing for knowledge of probabilistic causal invariance rather than deterministic disjunction provides a stronger test of our thesis.

# Preschooler Experiments

Our two studies with preschool children tested our causal-invariance hypothesis against alternative hypotheses, including ones in addition to the linear-integration rule tested in Liljeholm and Cheng (2007). The linear rule states that the observed value of the outcome is explained by the sum of the individual causal influences present. Our studies concern evaluating the effects of two treatments for removing (or preventing) an undesirable outcome, to decide which treatment best removes the outcome. Generalizing across contexts in the scenario involves generalizing from a farm context to a zoo context. Study 1 tested a situation in which the noisy-AND-NOT and linear integration rules yield opposite recommendations for action, and the divergence does not diminish with increased sample size.

Unlike the event frequencies in Liljeholm and Cheng's experiments, the event frequencies in Study 1 (see Table 1) were constructed so that logistic regression and the linear rule recommend the same action (see Cheng et al., 2003 for an explanation of the shared recommendation), contrary to that recommended by the noisy-AND-NOT rule. Logistic regression is a widely used statistical procedure in the medical sciences for evaluating the causal effects of treatments for binary outcome variables. Binary variables are common in medicine (e.g., whether or not a bone is fractured, a tumor is malignant, a woman is pregnant, a patient survives).

In both Studies 1 and 2, the children listened to an interactive story that concerns two brothers – a farmer and a zookeeper – who noticed that some of their animals had red dots on their faces. They were told, "The animals didn't seem sick at all, but the red dots made them look kind of funny." They heard that two "really tasty" and healthy treats, one a grain and the other leaves, might make the red dots go away. The brothers decided to figure out whether the treats work. First, they visited the farm, and fed the grain treat to every farm animal; later they visited the zoo, and fed both treats to every zoo animal.

Table 1 displays the pattern of event frequencies at the farm and at the zoo for Study 1. The critical "transfer" question is: To relieve red dots on new farm and zoo animals that have red dots on their faces, if one has to choose one and only one treat, what is one's best bet on which treat to use, grain or leaves? Assume that neither treat has any bad effects.

**Table 1: Event frequencies for Study 1**

| Intervention | **Farm** Grain only | **Zoo** Grain & leaves |
|---|---|---|
| Animals with dots: Pre intervention | 9/10 | 4/10 |
| Animals with dots: Post intervention | 6/10 | 1/10 |
| Number Cured | 3 | 3 |
| Fraction Cured | 3/9 | 3/4 |

Regardless of how "sameness of influence" is defined, the rationale underlying the choice is: Assuming the grain operates the *same* way across contexts (i.e., farm and zoo), then if the influence of the *intervention* (grain at farm vs. both treats at zoo) is observed to be the same across contexts, one's best guess would be that leaves had no influence – grain alone would already explain the outcome. But, if the influence of the intervention varied across contexts, one would attribute the difference to leaves.

Whereas the causal-invariance function predicts recommending leaves, models adopting a linear integration rule, frequentist or Bayesian, recommend using grain. Here we briefly sketch inferences according to the two rules (for prediction details see Cheng et al., 2013).

First, according to the noisy-AND-NOT integration rule (using Bayesian maximum-likelihood estimates of causal strengths as the predictor), the outcomes at the farm suggest that the grain removes red dots in a farm animal with a 1/3 probability. Assuming the grain's efficacy remains the same for the zoo animals as for the farm animals, grain would be expected to remove red dots with a probability of 1/3 in every zoo animal. It should be clear that the treat ingested by each animal does not "know" what the treat does in other animals (the independent-trials assumption). It follows that only 1/3 of the 4 zoo animals with red dots would be expected to have their red dots removed by the grain. Because in fact 3 of these 4 animals had their red dots removed, considerably more than 1/3 of the 4, the leaves must be explaining the large difference between the expected and the observed outcome. The causal-invariance function therefore predicts recommending leaves for the new animals. Note the use here of deviation from invariance as a criterion for revising one's causal beliefs, from grain as a preventive cause to leaves as a preventive cause also.

In contrast, according to the linear integration rule, because 3 of 10 animals were "cured" both at the farm and at the zoo, the addition of the leaves treat at the zoo does not result in any additional cured animals. This rule therefore predicts that the leaves treat is noncausal, and recommends giving the new animals grain.

The transfer question can be equivalently stated in terms of an interaction with something in the context. Both variants of the question address whether one's initial causal belief regarding relieving red dots requires revision.

## Study 1

### Method

**Participants** The participants were 29 children (13 male and 16 female) Children's mean age was 3.42 years (range 2.61 to 4.84 years, SD = .60 years). One additional child was excluded for failure to complete the task. Children were recruited from preschools in Los Angeles, CA. All children were fluent speakers of English and were learning English as a primary language.

**Procedure** As mentioned, children first listened to the story about the farm and zoo animals with and without red dots on their faces. The farm animals received a grain treat intervention and the zoo animals received a simultaneous grain and leaves treat intervention. In the last part of the study, children were shown new farm and zoo animals and asked to choose between two potential interventions.

**Storybook Task** The task was presented in a child friendly format, as an interactive storybook. The "reader" of the book was blind to any hypotheses of the study. Children were read the following cover story:

"Once upon a time there were two brothers, one was a farmer and the other a zookeeper. The two brothers loved their animals very much and took very good care of them. One day, the brothers noticed that some of their animals had red dots on their faces."

After being reassured that the animals were not sick, the children were told about the two treats, and were asked to determine their efficacy. They were told that both tasty treats would be loved by the animals.

"The two brothers decided to figure out whether the treats work. First, they went together to the farm. Then, they went over to the zoo. Let's look at what happened and see if YOU can figure out if the grain makes the red dots go away and if the leaves makes the red dots go away."

The farm context and the zoo context were presented separately, and the change in context was highlighted and emphasized. Animals in the farm context received the grain intervention only, whereas those in the zoo context received the grain and the leaves intervention in combination.

Figure 1 depicts examples of the pre- and post-intervention pictures that children saw. Because it was critical for children to attend to 1) the presence or absence of the red dots and 2) the administered intervention, those aspects of the story were interactive. For example, children were told "Here is a cow before it ate anything today" and then were asked "Does this cow have any red dots?" Children's responses were acknowledged (e.g., "You're right he does have red dots”). Children were then handed a cut-out of the treat to feed to the animal. Next, children were asked to make a prediction (e.g., “Do you think the cow will have red dots on its face now that it ate the grain?”) After the child replied, the experimenter said, “Let’s see!” and showed the picture of the treat inside the animal’s tummy, and the presence or absence of red dots was noted regardless of how the child answered (e.g., “Look no more red dots!”). This procedure was repeated with all twenty animals.

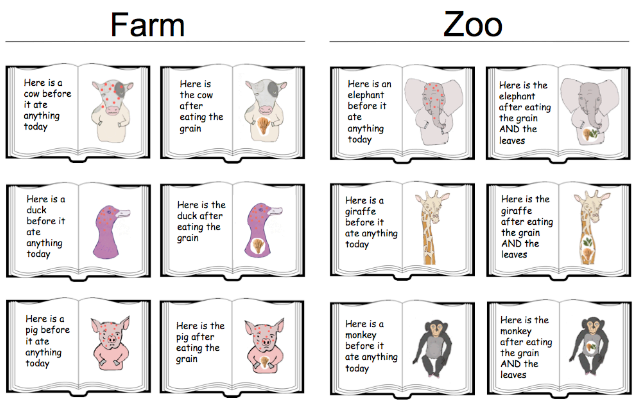


Figure 1: Examples of the pre- and post- intervention pictures.

**Treat Selection** The critical test was presented to children at the conclusion of the story. Children were shown two new animals (one farm and one zoo animal) with red dots on their faces and were asked to select only one of the treats, either the grain or the leaves, to make the animals’ red dots go away.

**Event Frequency** Table 1 depicts the event frequencies for the Study 1. To control for primacy and recency effects, the first trial at the farm and at the zoo showed the same event type; likewise, the last trial at the two locations showed the same event type. (A replication of the study randomized trial order; see note at end of Study 2.) As explained earlier, the noisy-AND-NOT integration rule predicts choosing leaves, but the linear integration rule predicts choosing grain. Note that the linear prediction requires a subset of the arithmetic steps required by the noisy-AND-NOT prediction. The linear rule also predicts the outcome at the zoo perfectly assuming fewer causes than the noisy-AND-NOT rule, namely, a single cause rather than two causes.

## Results

Children were attentive during the storybook reading and rarely responded incorrectly about the presence or absence of red dots. Across all children and all questions there were 7 initial incorrect responses (out of 360 total queries). For these seven responses children were corrected (e.g., “Look here are red dots”) and queried again.

The critical result concerned which treat children selected to make the animals’ red dots go away. As Figure 2 shows, children overwhelmingly chose the leaves $X^2$ (1) = 12.4, $p$ =.0004, suggesting that children’s responses fit with the noisy-AND-NOT rule rather than with the linear rule. They did so despite the linear rule’s relative arithmetic simplicity and its perfect accuracy predicting the outcome at the zoo using fewer causes.

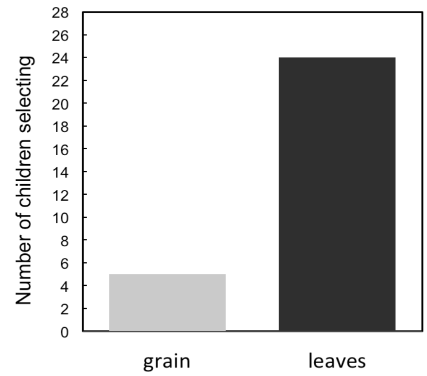


Figure 2: Results from Study 1 depicting the number of children selecting the grain treat versus the leaves treat.

# Study 2

There are alternative explanations for why the children selected leaves in Study 1. The children’s attention could be biased toward the newer second treat. The children might simply have a bias toward leaves. Or they might have used a heuristic: pick the treat uniquely associated with the *fewest* animals with red dots after the intervention. Previous related experiments have not ruled out analogous hypotheses. To rule out all three alternative explanations, Study 2 presented the same story but with the event frequencies in Table 2 to a separate group of preschoolers. As should be clear, the heuristics and biases still predict choosing leaves. For example, as before, fewer animals had red dots after the intervention at the zoo than at the farm (one and two, respectively). The noisy-AND-NOT rule predicts choosing grain this time; the “treatment” maintained the same preventive strength of ¾ at the farm and at the zoo. Along with the above heuristics and biases, the linear rule predicts no change from the recommended action in Study 1.

**Table 2. Event frequencies for Study 2**

| Intervention | Farm<br>Grain only | Zoo<br>Grain & leaves |
|---|---|---|
| Animals with dots: Pre intervention | 8/8 | 4/8 |
| Animals with dots: Post intervention | 2/8 | 1/8 |
| Number Cured | 6 | 3 |
| Fraction Cured | 6/8 | 3/4 |

## Method

**Participants** The participants were 28 preschool-aged children (M= 4.38 years, range 2.61 years – 5.18 years, SD = .66 years). 14 were male and 14 were female. An additional two children were excluded for failure to complete the task and/or attend to the story. Children were recruited similarly using the same criteria as for Study 1.

**Procedure** The procedure replicated that in Study 1 except that there were 16 trials in total, with the event frequencies for the farm and zoo animals as specified in Table 2.

**Results** As before, the critical result concerned which treat children selected to make the animals' red dots go away. Figure 3 shows that children's pattern of responses reversed in Study 2: children were now significantly more likely to select the grain treat, $X^2$ (1) = 5.14, $p$ =.02.

We replicated the pattern of results in Studies 1 and 2 in a variant in which the children were randomly assigned to the two studies, and the order of trials in each context (farm and zoo) was randomized for each child.

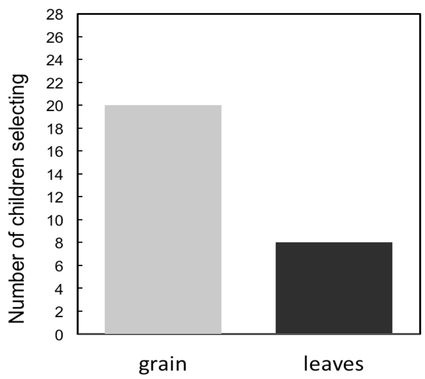


**Figure 3**: Results from Study 2 depicting the number of children selecting the grain treat vs. the leaves treat.

## Discussion

Our results favor young children's use of a causal-invariance function over use of the simpler linear function, a preference for one of the candidate causes, or a heuristic to choose the candidate more frequently paired with the desired outcome. Only the noisy-AND-NOT rule representing causal invariance can explain the opposite predominant choices across both our studies. More complex alternative hypotheses, such as use of the linear function in combination with a bias toward the candidate with the more frequent pairing, await further study.

The goal of our present paper is to provide support for the essentiality of the concept of causal invariance, as a default and a criterion for belief revision, in the construction of useable causal knowledge, when the set of possible causal representations is too large to exhaustively evaluate. Our findings indicating the early use of a probabilistic causal-invariance function -- embodying the rather abstract concept of the unchanging nature of the forces of change -- suggest that the generalizability of causal knowledge, along with parsimony and logical consistency, is not a mere wish but a constraint in the rational construction of causal knowledge.

## Acknowledgments


The research reported in this article was supported by AFOSR FA 9550-08-1-0489. We thank Chris Carroll and Reid Hastie for helpful discussions, and Maryann Francis and Natalie Peri for conducting the experiments.


## References


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

Cheng, P.W., Liljeholm, M. & Sandhofer, C. (2013). Logical consistency and objectivity in causal learning. In *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 2034-2039). Austin, TX: Cognitive Science Society.

Cheng, P.W. & Lu, H. (in press). Causal invariance as an essential constraint for creating a causal representation of the world: Generalizing the invariance of causal power. In M.R. Waldmann (Ed). *The Oxford Handbook of Causal Reasoning.* Oxford, England: Oxford Univ Press.

Gopnik, A. (2009). *The philosophical baby*. New York: Farrar, Straus and Giroux.

Gweon, H. and Schulz, L.E. (2011) 16-month-olds rationally infer causes of failed actions. *Science*, *332*, 1524.

Hume D. (1739/1987) *A treatise of human nature* (2nd edition, Clarendon Press, Oxford).

Legare, C.H. (2012) Exploring exploration: explaining inconsistent information guides hypothesis-testing behavior in young children. *Child Development*, *83*, 173–185.

Liljeholm, M., & Cheng, P.W. (2007). When is a cause the "same"? Coherent generalization across contexts. *Psychological Science*, *18*, 1014-1021.

Lu, H., Yuille, A., Liljeholm, M., Cheng, P.W., & Holyoak, K.J. (2008). Bayesian generic priors for causal learning. *Psychological Review*, *115*, 955-984.

Lu, H., Rojas, R. R., Beckers, T., & Yuille, A. L. (2016). A Bayesian theory of sequential causal learning and abstract transfer. *Cognitive Science, 40*, 404–439.

Lucas, C.G., & Griffiths, T.L. (2010). Learning the form of causal relationships using hierarchical Bayesian models. *Cognitive Science*, *34*(1), 113–147.

Lucas, C.G., Bridgers, S., Griffiths, T.L., Gopnik, A. (2014). When children are better (or at least more open-minded) learners than adults: Developmental differences in learning the forms of causal relationships. *Cognition*, *131*, 284-299.

Melchers, K. G., Lachnit, H., & Shanks, D. R. (2004). Past experience influences the processing of stimulus compounds in human Pavlovian conditioning. *Learning and Motivation*, *35*(3), 167-188.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Mateo, CA: Morgan Kaufmann.

Rakison, D. H. & Krogh, L. (2012). Does causal action facilitate causal perception in infants younger than 6 months of age? *Developmental Science*, *15*, 43-54.

Wheeler, D.S., Beckers, T. & Miller, R.R. (2008). The effect of subadditive pretraining on blocking: Limits on generalization. *Learning* & Behavior, *36* (*4*), 341-351.