PAPER

Blocking a redundant cue: what does it say about preschoolers' causal competence?

Heidi Kloos[1] and Vladimir M. Sloutsky[2]

1. Department of Psychology, University of Cincinnati, USA
2. Department of Psychology, The Ohio State University, USA

Abstract

The current study investigates the degree to which preschoolers can engage in causal inferences in a blocking paradigm, a paradigm in which a cue is consistently linked with a target, either alone (A-T) or paired with another cue (AB-T). Unlike previous blocking studies with preschoolers, we manipulated the causal structure of the events without changing the specific contingencies. In particular, cues were said to be either potential causes (prediction condition), or they were said to be potential effects (diagnosis condition). The causally appropriate inference is to block the redundant cue B when it is a potential cause of the target, but not when it is a potential effect. Findings show a stark difference in performance between preschoolers and adults: While adults blocked the redundant cue only in the prediction condition, children blocked the redundant cue indiscriminately across both conditions. Therefore, children, but not adults, ignored the causal structure of the events. These findings challenge a developmental account that attributes sophisticated machinery of causal reasoning to young children.

Introduction

Imagine a person is developing an allergic reaction after eating cereal with milk. Both cereal and milk could be potential causes for the allergy. Yet after positively identifying milk as a cause, one is likely to discard cereal as a potential cause. Discovering that milk caused the allergy appears to attenuate the possibility of cereal being a cause also. Now imagine that a person coughs and runs a fever. Both of these reactions could be potential effects of the flu. After positively identifying cough as the effect of flu, one is unlikely to discard fever as a potential effect. Discovering that the cough was caused by the flu does not attenuate the possibility of fever being the effect of the flu too. The two examples illustrate, on an intuitive level, that there is an asymmetry between how people think of multiple causes versus multiple effects: whereas causes are seen to compete, effects are not (Pearl, 1988, 2000; Reichenbach, 1956). The current study investigates whether preschoolers appreciate this asymmetry.

Understanding causality in multi-cue scenarios

There are many situations in which children, and even infants, seem to exhibit appreciation of causality (e.g. Bonawitz, Ferranti, Saxe, Gopnik, Meltzoff, Woodward & Schulz, 2010; Bullock, Gelman & Baillargeon; 1982; Gweon & Schulz, 2011; Leslie & Keeble, 1987; Sobel & Kirkham, 2006; Shultz & Altmann, 1982; Téglás, Vul, Girotto, Gonzalez, Tenenbaum & Bonatti, 2011): Preschool children can categorize novel objects on the basis of their causal property (e.g. Gopnik & Sobel, 2000); they can re-create or prevent an event (e.g. Gopnik, Sobel, Schulz & Glymour, 2001; Nazzi & Gopnik, 2003; Gopnik & Glymour, 2002; Gopnik & Tenenbaum, 2005, 2007; Glymour, 2003; Sobel, Tenenbaum & Gopnik, 2004); they can link causal properties with insides of objects (e.g. Buchanan & Sobel, 2008; Nazzi & Gopnik, 2000; Sobel & Munro, 2006, 2009); and they can select the correct cause when multiple candidate causes are available (e.g. Schulz & Gopnik, 2004). These findings appear to suggest that children are equipped with some

abstract causal knowledge, including 'assumptions about causal structure and about how patterns of events indicate causal structure' (Gopnik, Glymour, Sobel, Schulz, Kushnir & Danks, 2004, p. 6).

There are also many situations where children demonstrate a surprisingly fragile understanding of causal relations (see Bullock, 1985; Schlottmann, 1999). For example, 3- to 5-year-olds have difficulty distinguishing between causally relevant and causally irrelevant actions (e.g. Lyons, Young & Keil, 2007): Despite being asked to imitate only the causally relevant actions of the experimenter (e.g. lifting a lid to retrieve an object), preschoolers tended to imitate irrelevant actions as well (e.g. rotating and lowering a wand). Even in a highly controlled setting of objects being placed on a machine one by one (upon which the machine gets activated or not; Gopnik *et al.*, 2001), over half of the 3-year-olds attribute causal powers to objects that were shown *not* to activate the machine. In addition, irrelevant variations in the task context appear to strongly influence children's performance. For example, in a paradigm analogous to Gopnik *et al.*'s (2001), causal inference in 5- to 9-year-olds was disrupted when a 5-second delay was imposed (Siegler & Liebert, 1974; for similar effects with adults, see also Shanks, Pearson & Dickinson, 1989).

At best, existing findings on children's causal reasoning are mixed, showing evidence of competence in some conditions and lack thereof in other conditions. Adding to such inconsistencies is the fact that performance differs across tasks that have the exact same underlying causal structure. For example, children performed better when asked to determine an object's category membership (on the basis of known causal properties) than when asked to determine an object's causal properties (on the basis of known category membership), even though the exact same labels and the exact same causal properties were used (Gopnik & Sobel, 2000). Children also performed better when asked to link a causal property to the insides of an object than to the label of the object (Buchanan & Sobel, 2008). And they performed better when asked to link an object's insides to a label than to a causal property (Sobel, Yoachim, Gopnik, Meltzoff & Blumenthal, 2007). Furthermore, children's causal inference was more likely to occur when they were asked to create an event rather than to prevent it, even though the abstract causal relations were identical (Kushnir & Gopnik, 2007). Finally, children performed better when they had to infer a causal scenario from a target than predict a target from a causal structure (Kushnir & Gopnik, 2005; Schulz, Gopnik & Glymour, 2007).

Such asymmetries in performance undermine a simple claim of causal competence in young children. Given that children exhibit evidence of competence

under some conditions but not under others, we deemed it necessary to re-examine the fundamental premise of the causal competence hypotheses – children's appreciation of how events differ in their underlying causal structure.

Differentiation of underlying causal structures: evidence from adults

Contingencies between events can have one of various underlying causal structures, ranging from simple cause/effect relations to more complex multi-cause/multi-effect relations, inhibitory causes, spurious causes, causal chains, or circular causes (see Cheng & Novick, 1991; Waldmann, 1996). Importantly, contingencies often provide ambiguous information about the causal structure. Thus, a causal structure, a model or hypothesis of some sort, needs to be actively supplied by the perceiver, before contingencies can be interpreted in causal terms (see e.g. Tenenbaum & Griffiths, 2003; Tenenbaum, Kemp, Griffiths & Goodman, 2011; Sloman, 2005). One well-known paradigm to study whether humans impose such causal models on observed contingencies is the so-called blocking paradigm.

Commonly the blocking paradigm involves a two-step demonstration: In the first demonstration, cue A is consistently related to a target outcome (A-T), and in the second demonstration, an additional cue B is added to A (AB-T). This added cue B becomes redundant, compared to cue A, because it does not add any information regarding the presence or absence of the target T. A classic example of this paradigm, used with animals, involves the flashing of a light when a food pellet is given (A-T), and then adding a tone every time the light comes on (AB-T). Analogously, an example used with children involves an object consistently predicting the behavior of a box (A-T), after which the object is paired with a second object (AB-T). Indeed, animals, children, and adults presented with such demonstrations are more likely to link cue A than cue B with the target (e.g. Cobos, Lopez, Cano, Almaraz & Shanks, 2002; Dickinson & Burke, 1996; Kamin, 1969; Kruschke & Blair, 2000; Lopez, Shanks, Almaraz, & Fernandez, 1998; Sobel *et al.*, 2004; Williams, Sagness & McPhee, 1994). In other words, cue B is blocked from being linked to the target, even though it was consistently linked with the target (see also Larkin, Aitken & Dickinson, 1998).

Blocking the redundant cue from being seen as a cause of the target is appropriate: If one cue (A) can already account for the effect, a co-varying cue (B) is likely to be causally irrelevant. Recall one of the scenarios provided at the start of the paper: if two foods are ingested at the

same time, followed by an allergic reaction, and one of the two foods has been established as causally relevant, the other food can be disregarded as having caused the allergy. However, even though blocking of the redundant cue is compatible with causal understanding, it does not necessarily imply such understanding. This is because blocking can be explained by purely associative models (e.g. a variant of the Rescorla-Wagner model proposed by Wasserman and colleagues, see Van Hamme & Wasserman, 1994; Wasserman & Berglan, 1998). In fact, blocking a redundant cue is a default that occurs in the absence of a causal scenario (see Hall, 1991, for a review of blocking as a consequences of latent inhibition in the associative process).

Therefore, in order to establish causal competence, it is not sufficient to demonstrate blocking performance. One would have to demonstrate – in addition – that blocking takes place only when it is causally appropriate, namely when two co-varying cues are potential causes of an outcome, not when they are potential effects of T (see Pearl, 1988, 2000; Reichenbach, 1956). Recall the second scenario we presented above: If two symptoms appear at the same time, and one of them has been established as the effect of a disease, the other one does not need to be disregarded as effect. To establish causal competence, one would have to demonstrate the ability to block the redundant cue selectively, namely in the case of multiple potential causes, but not in the case of multiple potential effects.

In an influential study, Waldmann and Holyoak (1992) demonstrated that adults appreciate this asymmetry between multiple-cause and multiple-effect events. Adults were presented with the two-phase contingencies A-T; C-¬T and AB-T; C-¬T (where ¬T stands for non-target). In the *Prediction* condition, they were told that A, B, and C are potential causes of T, and the task was to predict whether a cue causes T or ¬T. Conversely, in the *Diagnosis* condition, participants were told that A, B, and C are potential effects of T, and the task was to diagnose whether a cue was caused by T or ¬T. The argument was that participants in the Prediction condition construe the observed contingencies as AB → T and A → T, and – as a result – block the redundant cue B from being associated with T. Participants in the Diagnosis condition, on the other hand, were assumed to construe the observed contingencies as AB ← T and A ← T, and thus do not block the redundant cue B. Findings support this argument: cue B was more likely to be blocked in the Prediction condition than the Diagnosis condition (see also Ali, Chater & Oaksford, 2011, for related arguments). In the current study, we examine whether young children also appreciate this asymmetry.

Experiment 1

Preschool children and adults first learned that event A was linked to target T (A-T), and that event C was linked to non-target ¬T (C-¬T). Then they learned that the pair of events A and B was also linked to target T (AB-T). A and B were either described as potential causes of T (*Prediction* condition) or as potential effects of T (*Diagnosis* condition). The crucial task was to either 'predict' whether B alone would cause T (Prediction condition) or to 'diagnose' whether B alone was caused by T (Diagnostic condition). The expectation was that a causally sophisticated learner appreciates the fact that multiple causes are more likely to compete with each other than multiple effects (see Sobel *et al.*, 2004, for related arguments). Therefore, if children are sensitive to the causal structure of the situation, they should be more likely to block the redundant cue B when it is introduced as a potential cause (as shown in Figure 1a) than when it is introduced as a potential effect (as shown in Figure 1b). In contrast, if children are not sensitive to the underlying causal structure, but nevertheless understand the task instructions, they should exhibit equivalent blocking in both conditions.

Method

Participants

Participants were 33 4- and 5-year-olds (16 girls and 17 boys), recruited from suburban middle-class preschools, and 27 introductory psychology students who participated for class credit. Half of the participants in each age group were randomly assigned to the Prediction condition, and the other half were assigned to the Diagnosis condition. The mean age for children was 61.3 months in the Prediction condition ($SD = 3.1$) and 59.3 months in the Diagnosis condition ($SD = 4.9$). Thirteen additional
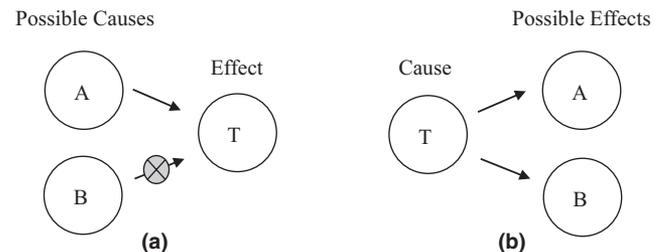
Figure 1 *Schematic representation of causal graphs to reflect the contingencies of the blocking paradigm A-T, AB-T. (a) A and B are said to be possible causes of T, with the expectation that the B-T connection is blocked. (b) A and B are said to be potential effects of T, with the expectation that the B-T is not blocked.*

Table 1   *Example stimuli of cues and pictorial representation of cover story used in Experiment 1*

	Stimuli		
Event Cues (examples)	A	C	AB
Target Cues	T:		¬T:

	Cover Story
Prediction Condition: Eating from the magic fruit turns the creature into a flower	
Diagnosis Condition Eating from the magic flower turns the creature into a plate of fruits	

children and three adults were tested and omitted from the sample because (1) they chose the 'impossible to tell' choice across all testing trials (one child from the Prediction condition, and two children from the Diagnosis condition), they did not meet the learning criterion on check trials (four children and two adults from the Prediction condition, and three children and one adult from the Diagnosis condition; see Procedure), or they did not reproduce the cover story at the end of the experiment (three children from the Diagnosis condition).

Materials

Materials were identical for both conditions, and they were presented on a laptop computer. The top part of Table 1 shows example stimuli of this experiment. In particular, the cues A, B, and C were realistic pictures of fruits (e.g. apple, orange, or a strawberry) measuring about 5 cm in diameter, and presented on a cartoon plate measuring about 12 cm in diameter. Each fruit was always displayed on one half of the plate. Depending on the trial, the plate had either one fruit or two. In the former case, only one fruit was depicted on the plate, and a napkin covered the rest of the plate.

The target cue T was a picture of an unusual-looking green-and-white flower. And the non-target ¬T was a picture of a stick figure scratching its head to signify 'impossible to tell'. This answer option, although different from what was done before, approximates the logically correct 'impossible to tell' answer when a cue is blocked, and thus encourages above-chance accuracy for participants who block a cue.[1]

Procedure

For this and all subsequent experiments, adults were tested in the lab on campus, and children were tested in a quiet room at their preschool. All experiments were administered on a computer and controlled by Super-Lab Pro 2.0 software. The cover story involved Toto, the traveler, who wants to find out about creatures in a far-away place. The bottom part of Table 1 shows how the cover stories differed between conditions. In the Prediction condition, fruits were said to be potentially magic. If a creature eats from magic fruits, it turns into the green-and-white flower. And if it eats from other fruits, is it impossible to tell what the creature will turn into. Toto's task is to predict what the creature will turn into after eating from a particular

---

[1] Previous research conceptualized the non-target as the opposite of the target (e.g. not-flower). This choice predicts chance performance for participants who block a cue, because the blocked cue is consistently associated neither with the target nor with the non-target.

Table 2 *Schematic illustration of the blocking-paradigm manipulation*

Training Trials	Block 1	A – T
		C – ¬T
	Block 2	AB – T
		C – ¬T
Testing Trials	Check	A ? (T or ¬T)
		C ? (T or ¬T)
		AB ? (T or ¬T)
	B	B ? (T or ¬T)

Note: A, B, C, T, and nonT represent cues, and the dash '–' signifies the cue contingency.

plate (or whether it is impossible to tell). In the Diagnosis condition, in contrast, plates of fruits were said to be what creatures turn into. A creature turns into a particular plate of fruits after eating from enchanted plants. If they turn into a different plate of fruits, it is impossible to tell what they ate. Toto's task in this case is to determine what the creature ate before it turned into a particular plate (or whether it is impossible to tell).

A warm-up phase was administered to ensure that participants understood the meaning of 'impossible to tell'. Participants were shown a picture of the napkin and asked to guess what it might cover. They were told, for example: 'Is there a doll underneath the napkin, or is it impossible to tell what's underneath it?' Feedback was given (the correct answer was always 'impossible to tell'), and a similar question was asked until participants performed correctly on two consecutive trials.

The experiment consisted of a training phase and a testing phase, administered back-to-back. During a training trial, a plate appeared on the screen, followed either by the flower (on T trials) or the 'impossible to tell' stick figure (on ¬T trials). The experimenter repeated the information about the causal status of the cues during each trial. For example, participants in the Prediction condition heard: 'What will the creature turn into if they eat from this plate? Look, it turns into the flower.' And participants in the Diagnosis condition heard: 'What did the creature eat before it turned into this plate? Look, it ate the flower.'

Table 2 shows the type of training and testing trials in schematic form (identical across conditions). During the first training block, cue A was consistently linked to the flower (A-T), while cue C was consistently linked to 'impossible to tell' (C-¬T). There were 15 A-T trials and 10 C-¬T trials, presented in random order. The high number of trials during this first block was used to ensure that participants could learn the target associations (see Waldmann & Holyaok, 1992, for a similar procedure). During the second training block, cues A and B appeared together on a plate and were consistently

linked to the flower (AB-T). The control cue C was again consistently linked to 'impossible to tell' (C-¬T). There were five AB-T trials, and five C-¬T trials, presented in random order.

There were two kinds of testing trials: (1) check trials – those that measured learning of contingencies (A-T, AB-T, C-¬T), and (2) critical trials – those that measured inferences about whether cue B is linked to T. During each testing trial, participants were shown a plate and asked whether it was linked to cue T or to cue ¬T. Specifically, participants in the Prediction condition heard: 'What will the creature turn into if it eats from this plate? Will it turn into the enchanted flower or is it impossible to tell?' And participants in the Diagnosis condition heard: 'What did the creature eat that turned into this plate? Did it eat from the enchanted flower or is it impossible to tell?' No feedback was given throughout testing. There were five check trials of each kind (A, C, AB) and 10 critical trials (B), presented in random order. To be included in the sample, participants had to perform correctly on at least 70% of check trials (i.e. A, AB, and C trials combined). Excluded participants constituted 15.2% of all tested children and 10% of all tested adults.

At the end of the experiment, children were asked to reproduce the story line of the experiment. In particular, they were shown the set of pictures displayed in the bottom part of Table 1 and were asked to choose (1) whether creatures ate a plate of fruit and then turned into a flower, or (2) whether creatures ate the flower and then turned into a plate of fruit. To be included in the sample, children had to perform correctly on this question. Of all tested children, 6.5% of children did not pass this criterion.

Results and discussion

A preliminary analysis pertained to participants' accuracy during check trials, when they were asked whether cues A, C, and AB were linked to T or ¬T. Figure 2A shows the mean proportion of correct answers on these trials, separated by age group (adults vs. children) and condition (Prediction vs. Diagnosis). Participants performed virtually at ceiling ($M = .95$), with no effects of age, condition, or trial type, and no significant interactions, mixed-design ANOVA, $Fs < 1.2$, $ps > 0.7$.[2] Finding such high performance on check trials shows that both children and adults could learn the cue-target relations presented during training.

[2] The pattern of results stayed the same when we included participants who did not pass the learning criterion of 70% (seven children and three adults). In this case, average performance on check trials was $M = .89$.
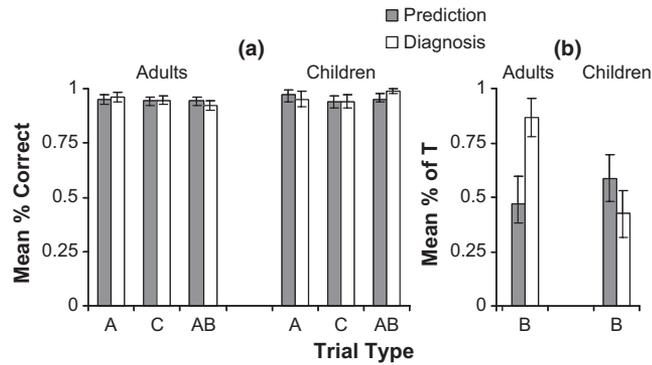
Figure 2 *Mean performance in Experiment 1 as a function of condition and age group. Panel A: Performance on check trials A, C, and AB, represented as the proportion of correct answers. Panel B: Performance on critical (or B) trials, represented as the proportion of T answers (vs. ¬ T answers). Error bars represent standard errors.*

How did children perform on critical test trials, when asked about the relation between cue B and T? Figure 2B shows the mean proportion of T answers on B trials, separated by age group and condition. To determine the degree of blocking, we conducted two separate mixed-design 2×2 ANOVAs, one for adults and one for children, using trial type (A vs. B) as the within-group factor, and condition (Prediction vs. Diagnosis) as the between-group factor. As is evident in a significant cue $\times$ condition interaction, $F(1, 25) = 6.0$, $p < .03$, adults were more likely to block cue B in the Prediction condition when B was said to be a potential cause of T ($M_{BT} = 0.47$ vs. $M_{AT} = 0.95$) than in the Diagnosis condition, when B was said to be a potential effect of T ($M_{BT} = 0.87$ vs. $M_{AT} = 0.96$). Children, on the other hand, blocked cue B both in the Prediction condition ($M_{BT} = 0.59$ vs. $M_{AT} = .98$) and in the Diagnosis condition ($M_{BT} = 0.42$ vs. $M_{AT} = .95$), as is evident in a main effect of trial type, $F(1, 31) = 7.2$, $p < .01$, and a missing effect of condition or condition interaction, $Fs < 1.2$.

Note that performance on B trials was at chance, for children and adults in the Prediction condition, as well as for children in the Diagnosis condition, single-sample $ts < 1$ (assuming a chance probability of 0.5). While this finding is consistent with previous research (e.g. Sobel *et al*., 2004[3]), it is nevertheless surprising here, given our use of the 'impossible to tell' answer option. We

[3] The participants in Sobel *et al*. (2004) were younger than in the current experiment (3- and 4-year-olds), and the tasks were somewhat different. Nevertheless, blocking performance is quite similar: In a categorization task, the redundant cue B was linked to the target in 50% of the cases.

predicted that a blocked cue would yield a large proportion of 'impossible to tell' choices, rather than a mix of both answer options. It is possible that participants tended to avoid committing to indeterminacy (see Beck, Robinson, Carroll & Apperly, 2006; Hoemann & Ross, 1971; Kuzmak & Gelman, 1986), despite having passed the initial warm-up phase (see Procedure). We nevertheless used the same answer option for all subsequent experiments, to ensure comparability across experiments.

Taken together, several findings emerged. First we were able to replicate previous findings with adults, using a novel task context: Adults blocked the redundant cue B more when it was presented as a potential cause of T than when it was presented as potential effect of T (see Waldmann & Holyoak, 1992). Second we were able to replicate previous findings with preschoolers exposed to a multiple-causes scenario: preschoolers blocked the redundant cue B when it was said to be a potential cause of T (see Sobel *et al*., 2004). Both of these findings establish that the context chosen for the current experiments is compatible with the previously chosen contexts. Third, children exhibited near-adult performance on check trials, indicating that children had little difficulty learning the contingencies. And finally, preschoolers blocked the redundant cue B regardless of whether it was a potential cause or a potential effect. Therefore, whereas both children and adults were sensitive to contingency information (as evidenced by the ability to associate cues and targets), only adults exhibited appreciation of the underlying causal structure (as evidenced by the asymmetric blocking in the Prediction and Diagnosis conditions).

This latter finding suggests that children do not differentiate between predictive and diagnostic conditions in the ways that adults do. While this finding is important, the current experiment does not eliminate several alternative explanations. First, it is possible that children failed to construe the evidence as causal because the cover stories were confusing and lacked mechanistic information (see e.g. Ahn, Gelman, Amsterlaw, Hohenstein & Kalish, 2000; Sobel & Munro, 2009). Or children might have difficulty specifically with the diagnostic context, which might arguably be a more unusual inference than a prediction (see Fernbach, Darlow & Sloman, 2010, 2011). We sought to address these issues in the subsequent experiments.

Experiment 2

Could the patterns of results found in Experiment 1 reflect some unrelated differences between children and adults? For example, could it be that the context was too

complicated and confusing, thus resulting in children's failure to appreciate the difference between the Prediction and the Diagnosis conditions? To discount these possibilities, we set out to create a causal context that was intended to be explicitly confusing (while keeping the cue–target contingencies the same): On some trials, cues were said to be the potential causes of the target, and on other trials, the same cues were said to be potential effects. We expected that such switching of causal scenarios would make it difficult to apply a causal framework to the evidence. Thus any similarity between the performance patterns obtained in this condition and that obtained in Experiment 1 would reflect the degree to which the latter patterns were due to confusion about the causal context.

Method

Participants

Participants were 20 4- and 5-year-olds (14 girls and six boys; $M = 64.2$ months, $SD = 5.3$), none of whom had participated in the previous experiment.

Materials and procedure

Materials and procedure were identical to those used in Experiment 1, and the cover story involved the same fictitious creatures that were used in Experiment 1. However, the causal status of the cues was manipulated across trials. In particular, children were told on some trials that creatures would turn into a plate of fruits after eating from the magic flower, and on other trials they were told that creatures would turn into the flower after eating from magic fruits. For every plate of fruit presented on the screen, the experimenter informed the child whether it showed what a creature had turned into, or whether it showed what a creature ate from (i.e. a small circle appeared either on the top left or the top right of the screen to signal to the experimenter whether the plate of fruits being shown should be presented as a cause or as an effect).

Results and discussion

A preliminary analysis looked at the degree to which children could learn the contingencies presented to them (i.e. proportion of correct responses on check trials A, C, and AB). Recall that in Experiment 1, we applied a learning criterion (i.e. to be at least 70% correct on check trials), which 15.2% of the children failed to meet in that experiment. In the current experiment, many more children did not meet this criterion (11 out of 20, 55%;

Fisher's Exact Probability, comparing the distribution of children in Experiment 1 vs. 2, $p = .02$). This finding is a first indication that many more children were confused about the overall task context in this experiment than in Experiment 1.

Looking at overall performance on check trials, children's proportion correct in the current experiment averaged .67, .75, and .71 for A, C, and AB trials, respectively. By comparison, overall performance on check trials in Experiment 1 (without applying a learning criterion, and collapsed across conditions, given that there was no difference between conditions) was .92, .91, and .86, for A, C, and AB trials, respectively. Confirming this difference between experiments, a 2 (experiment) by 3 (trial type) mixed design ANOVA revealed a significant effect of experiment, $F(1, 54) = 5.7$, $p < .05$. There was no effect of trial type nor a trial-type interaction.

Did children block the redundant cue when they were presented with a confusing task context and had difficulty learning the cue–target contingencies? Given that children's learning of the cue–target contingencies was far from ceiling, the degree to which children associated cue B with target T was measured in relation to the degree to which they associated cue A with target T (rather than in absolute terms as average proportion of BT associations). For each child, we therefore determined the difference between the proportion of AT associations and the proportion of BT associations. A positive difference implies blocking of the redundant cue B, while no difference (or a negative difference) implies that the redundant cue B was linked to target T to the same degree as (or more than) cue A.

The average of these difference scores is shown in Figure 3. As can be seen in the figure, when we included all participants of this experiment, the mean AT – BT difference was zero. By contrast, the mean AT – BT difference in Experiment 1 is far from zero ($M = .46$), single-sample $t(32) = 4.48$, $p < .01$.[4] Even if we apply the exact same learning criterion and include only children who obtained a learning score of 70% or higher ($N = 9$ out of 20), difference scores nevertheless differed between the two experiments, independent-sample $t(40) = 2.01$, $p < .04$.

Taken together, children's performance on B trials of the current experiment did not show the signature blocking pattern that was observed in Experiment 1. Thus, when children participated in a task context that was explicitly confusing, they did not block cue B. This

[4] Here we collapsed across Prediction and Diagnosis conditions because there was no difference between the two. The same pattern of results holds when we include only one of the conditions.
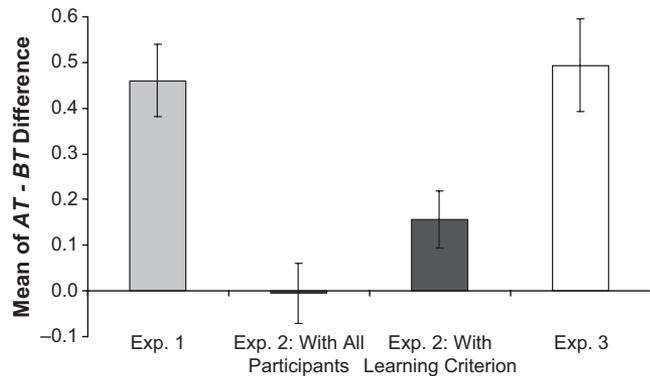
Figure 3 *Children's mean difference scores (proportion of AT associations minus proportion of BT associations) in Experiments 1–3. For Experiment 2, data pertain to whether all participants are included (first bar) or only those participants who passed the learning criterion applied for Experiments 1 and 3. Error bars represent standard errors.*

held true even when we omitted children who did not pass the learning criterion used in Experiment 1, discounting the possibility that children's performance in Experiment 1 reflects confusion on the part of the child. In the next experiment we provide further evidence for this claim, this time using a task context that was non-causal and thus omitted a cover story that could potentially confuse children.

Experiment 3

Children were again presented with the same contingencies (see Table 2), but no causal story was presented at all. Children were merely asked to learn the cue–target contingencies. Removing the causal cover story of magical transformations has two implications: First it reduces possible confusions about the cue–target contingencies. Children's performance is therefore likely to reflect some form of pure patterns of contingency learning. And second, it does not explicitly invite children to apply their causal knowledge. Performance might therefore reflect general learning processes, unrelated to a specifically predictive or diagnosis context.

Method

Participants

Participants were 16 4- and 5-year-olds (five girls and 11 boys; $M = 61.8$ months, $SD = 3.5$), none of whom had participated in the previous experiments. Four additional children were tested but were omitted from the sample because of (1) choosing the 'impossible to tell' choice

across all testing trials (one child), or (2) failing to meet the learning criterion on check trials (three children; see Procedure in Experiment 1).

Materials and procedure

Materials and procedure were identical to those used in Experiment 1, with the exception of omitting the causal cover story of magical transformation. Children were merely told that cues 'go together' with the target. Their task was to find out 'which plate of fruits goes with the flower, and for which plate of fruits is it impossible to tell what it goes with?' Using the same learning criterion applied in Experiment 1, participants had to perform correctly on at least 70% of check trials (i.e. A, AB, and C trials combined).

Results and discussion

As was done in Experiment 1, children's performance on check trials (A, C, and AB) was represented as the proportion of correct responses, whereas children's performance on critical B trials was represented as a proportion of T choices. Considering performance on check trials only, a 2 by 3 mixed-design ANOVA revealed no effect of experiment (Experiment 1 vs. Experiment 3), no effect of trial type (A, C, or AB), and no interaction between experiment and trial type, $Fs < 1$.[5] That is to say, children could learn the cue–target contingencies equally well, independently of whether they were presented with a causal story (Experiment 1: $M = .96$) or not (Experiment 3: $M = .95$).

Similarly, when considering only B-trial performance, there was no difference between the proportion of BT associations between Experiment 1 ($M = .50$) and Experiment 3 ($M = .47$), independent-sample $t < 1$. Figure 3 shows the mean difference between AT and BT associations. Again, there was no difference between Experiments 1 and 3, $t < 1$, suggesting that children blocked cue B to the same degree, whether they were explicitly provided with information about the causal status of cues or whether this information was omitted.

Experiment 4

Could children's performance in the Diagnosis condition of Experiment 1 stem from their having been asked to

---

[5] Data from Experiment 1 were again collapsed across Prediction and Diagnosis conditions, given that there was no difference between the two conditions. The same patterns of results were found when the analyses were conducted for each condition separately.

engage in a relatively unusual inference? In Experiment 4, we examined this question by using a new set of stimuli and a new causal scenario, potentially simpler and more familiar to young children. Rather than invoking the magical transformation of creatures, we used a scenario in which creatures leave behind foot-prints and other marks. In the Diagnosis condition, children had to determine whether the presence of a mark (effect) indicates the presence of a certain creature (cause). This is analogous to determining the animal by its footprints – a task context that should be rather familiar to children. In the Prediction condition, on the other hand, children had to predict whether the presence of a predator's mark (cause) would scare away prey (effect). This is analogous to predicting that a mouse would be scared upon seeing the paw prints of a cat.

Method

Participants

Participants were 25 4- and 5-year-olds (12 girls and 13 boys) randomly assigned to one of the two conditions (Prediction vs. Diagnosis). None of them had partici-pated in the previous experiments. The mean age was 60.1 months for children in the Prediction condition (SD = 2.1) and 57.6 months for the Diagnosis condition (SD = 2.0). An additional 10 children were tested and omitted from the sample (six from the Prediction condition and four from the Diagnosis condition) because they did not meet the learning criterion on check trials (see Procedure of Experiment 1). No child chose the 'impossible to tell' choice across all testing trials.

Materials and procedure

The top part of Table 3 shows examples of cues used in this experiment. Cues A, B, and C were colorful drawings of marks of animals, selected from the follow-ing set: a hairball (shown as cue A in Table 3), a footprint (shown as cue B in Table 3), a scratch made by a sharp tooth (shown as cue B in Table 3), and a bite mark on a log (shown in the cover-story display in the bottom part of Table 3). The target cue, referred to as the creature, was a grey amorphous shape with eyes. And the non-target ¬T was the same stick figure used in Experiments 1–3 to represent 'impossible to tell'.

Training and testing were identical to the procedure used in Experiment 1 (see Table 2), differences pertain-ing only to the differences in instructions and explana-tions that corresponded to the new stimuli. In particular, children were told that Toto wants to find out about a novel creature that lives in a far-away place. Toto traveled to that place and took pictures of animal marks he found there. Children in the Diagnosis condition were told that marks came either from the creature or from other animals. The task was to learn whether a mark was made by the creature. In the Prediction condition, children were told that marks come either from the predator of the creature or from other animals. The creature gets scared and runs away when it sees the marks of its predator, but not when it sees marks of other animals. The task was to find out whether a mark would cause the creature to get scared or not. The bottom part of Table 3 shows the pictorial representation of the cover stories used in the experiment.
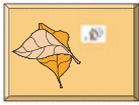
During familiarization, children were presented with a picture of a feather, a bird's foot print, a cat's paw print, and a horseshoe. Each mark was explained verbally (e.g. 'Look. This is a feather'). Children in the Diagnosis condition were asked for each mark: 'Was it a bird that left this mark behind?' Feedback was provided until children performed correctly across all four trials (responding 'yes' to the feather and the bird's footprint). And children in the Prediction condition were told: 'A cat eats mice. If a mouse comes along and sees this mark, will it be scared and run away?' Feedback was again provided until children performed correctly across all four trials (responding 'yes' to the cat's paw print). All children performed correctly after no more than one prompt.
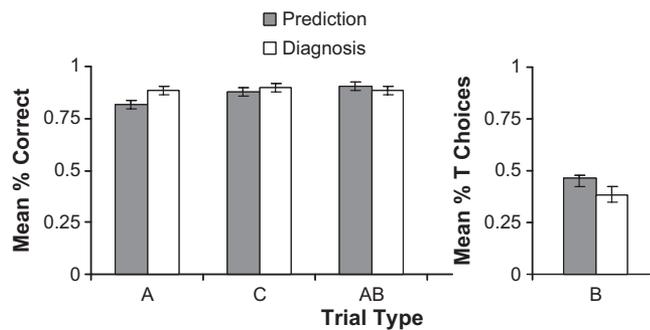
Results and discussion

Figure 4 shows children's performance as a function of trial type (check trials: A, C, AB; critical trials: B) and condition (Prediction vs. Diagnosis). The first analysis pertains again to children's performance on check trials (measured as proportion correct). As can be seen in the figure, performance on check trials was quite high, independently of condition or trial type, 2 by 3 mixed-design ANOVA Fs < 1.3. Similar to what was found in Experiment 1, children could learn the cue–target contingencies presented to them (M = 0.88).

To assess the degree of blocking the redundant cue B, a mixed-design ANOVA was conducted, with experiment (Experiments 1 vs. 4) and condition (Prediction vs. Diagnosis) as the between-group factors, and trials type (AT vs. BT) as the within-group factor. We again found only an effect of trial type, with no difference between experiments or conditions, and with no two-way or three-way interaction, Fs < 1.1. This implies that chil-dren's blocking was unaffected by a predictive versus diagnostic causal context: While children associated A with T to a high degree (M = .85), they did so far less for

Table 3 *Example stimuli of cues and pictorial representation of cover story used in Experiment 4*

	Stimuli

Event Cues (examples)   A   C   AB

Target Cues   T: ¬T:

	Cover Story

Prediction Condition: A mark made by his enemy will scare the creature away.

Diagnosis Condition The creature makes some marks

Figure 4 *Children's mean performance as a function of trial type (Check Trials vs. B trials) and condition (Prediction vs. Diagnosis) in Experiment 4. Performance on check trials is represented as proportion correct, whereas performance on B trials is represented as proportion of T choices. Error bars represent standard errors of the mean.*

the BT association ($M = .42$), paired-sample ts > 4.2, ps $< .001$. That is to say, the change in content from Experiment 1 to the current experiment (magical transformations vs. creature presence) did not affect overall patterns of performance.

Taken together, the findings of Experiment 1 were replicated with a new, conceivably simpler task context.

Specifically, the current findings establish the generality of the findings of Experiment 1, ruling out that they were driven by idiosyncratic details of magical transformations. Note also that the degree of blocking in Experiments 1 and 4 was equivalent to preschoolers' blocking in the no-cause condition of Experiment 3, when the cue–target links were explicitly non-causal.

General discussion

The goal of the current paper was to determine whether preschool children could distinguish between multiple-cause and multiple-effect scenarios when asked to evaluate the causal status of a redundant cue. Formally, multiple cues compete when they are candidate causes, but not when they are candidate effects. Findings show a stark difference between preschoolers and adults in terms of whether they can appreciate this distinction: While adults blocked the redundant cue only in the multi-cause scenario, children blocked it in both conditions, even in the multi-effect scenario. In other words, while adults were affected by the underlying causal structure of the events conveyed in the cover story, children were not.

Can preschoolers distinguish between multi-cause and multi-effect scenarios?

There are two lines of argument that could undermine the conclusion that preschoolers have difficulty distinguishing between multi-cause and multi-effect scenario. One line of argument centers on postulating confusion on the part of the young child (the confusion making it difficult for the child to display their full potential). And the second line of argument centers on postulating early causal competence that is locally mistaken (yielding mistaken performance despite early competence). We discuss each possibility in turn.

In terms of the first line of argument, there are indeed several reasons why children could be confused about the task. For example, confusion could be due to limited attentional capabilities required to keep track of the fairly complex cover story (see e.g. Frye, Zelazo, Brooks & Samuels, 1996; Zelazo, Reznick & Piñon, 1995). Or children have difficulty understanding the indeterminacy of the 'impossible to tell' answer option (see e.g. Beck *et al.*, 2006; Hoemann & Ross, 1971; Kuzmak & Gelman, 1986). Furthermore, it is possible that children have difficulty reasoning backwards from effects to their cause (see e.g. Fernbach *et al.*, 2010, 2011), which was necessary in the Diagnosis condition. Or children could be confused when a causal scenario lacks mechanistic information (see e.g. Ahn *et al.*, 2000; Sobel & Munro, 2009). Of course, one could also assume confusion to result from any combination of these factors. For example, young children may find it difficult to integrate the linguistic cues with the cue–outcome events presented to them, especially given the conflict between linguistic cues and children's assumption that causes appear before effects. In other words, confusion might stem from requiring children to form an integrated representation of current verbal descriptions and prior associative knowledge.

Is it possible that children's indiscriminant blocking in both the prediction and diagnosis condition could be attributed to their confusion about the task? Although it cannot be ruled out completely, there are several findings that speak against this suggestion. First, children ably learned the contingencies presented to them when the causal structure stayed unchanged (Experiments 1, 3 and 4), but showed markedly lower performance when the causal structure was explicitly designed to be confusing (Experiment 2). Second, learning of contingencies was equally high (and at the level of adult learning) whether the contingency pertained to the target (A-T; AB-T) or the

non-target 'impossible to tell' (C-¬T).[6] Third, their learning of contingencies did not differ as a function of predictive versus diagnostic scenarios, whether the diagnostic context was likely to be familiar to children (guessing the animal that left behind a track; Experiment 4) or not (guessing what a creature ate before turning into something else; Experiment 1). Finally, children could remember the causal structure of the events, as assessed in the memory trials at the end of the experimental session. Finding similarly high performance on check trials, despite differences in age, trial type, causality of cue–target relations, or details of the cover story, implies that children could understand the cover story and task instructions sufficiently well to apply causal-reasoning strategies.

In terms of the second line of argument, children's performance might stem, not from difficulty with encoding the cause–effect relations presented to them, but from holding mistaken ideas about these cause–effect relations. Locally mistaken causal knowledge might pertain to over-applying a predictive causal structure, for example to a scenario that has a diagnostic causal structure or no causal structure at all. Or it might pertain to assuming that potential effects compete (in the same way that potential causes compete). The idea is supported by findings showing that even adults, although highly competent causal reasoners, nevertheless fail to conform to formal causality in some task contexts (see e.g. Ali *et al.*, 2011; for a review see De Houwer & Beckers, 2002).

Is it possible that children misinterpreted the contingencies as multi-cause events even in the case in which the causal context did not support such a construal? If so, it would be formally appropriate to block the redundant cue, and thus confirm causal competence. We argue against this possibility, for several reasons. Consider first Experiment 3, where the contingencies between plates of fruits and the flower had no causal cover story. It is highly implausible that young children would supply such causal relations spontaneously, given that there is no transparently mechanistic link between fruits and flowers. Furthermore, it is not clear why children would prefer to construe a multi-cause predictive context over a multi-effect diagnostic context. In Experiment 4, the diagnostic scenario is likely to be more familiar to young children (being asked to identify the animal that left behind a particular track) than the predictive scenario (being asked to identify the track that would scare an animal). Indeed, anecdotal evidence from the preliminary training on the causal structure of the scenarios suggests more difficulty in the predictive than the diagnostic scenario of this experiment.

[6] Recall also that children succeeded in a preliminary feedback training during which they had to use the 'impossible-to-tell' answer option.

Finally, is it possible that children, while capable of distinguishing between multi-cause and multi-effect scenarios, mistakenly believe that multiple effects compete? Without having explicitly probed children's knowledge about cue competition, one can only speculate about whether blocking performance in the diagnosis condition is the result of mistaken causal knowledge. We argue that children's experiences are unlikely to justify a belief that potential effects compete. Their own actions are likely to have more than one effect (pushing a glass off the table causes a loud noise *and* a broken glass), and in fact, one would be hard pressed to find any examples of competing effects in the same way that all caused events have more than one feature. Looking specifically at the current experimental context, there was nothing in the diagnosis conditions that would imply a competition among possible effects. In Experiment 1, creatures were said to turn into entire plates of fruits (not just a single fruit), and in Experiment 4, creatures could leave behind more than one mark (e.g. the familiarization phase included two marks of a bird). Thus, while children might have the wrong ideas about multiple effects (vs. about multiple causes), there is little in their everyday experience or in the current experiment that could warrant such ideas.

Taken together, we argue that the pattern of performance of preschoolers in predictive versus diagnosis scenarios cannot be dismissed on the basis of superficial or idiosyncratic features of the task. There is little in the patterns of performance that would indicate confusion on the part of the child; and it is unlikely that children merely applied the wrong causal ideas to the scenarios. Instead, our findings are likely to add to the discussion of the processes that guide children's causal reasoning, with no more burden of proof than findings of early success. In the next section we elaborate on this discussion.

What are the processes that guide early causal reasoning?

The question of what guides causal thought in young children has a long tradition, with a multitude of experimental methods, ranging from perceptual habituation paradigms to active predictions and problem solving, yielding a multitude of findings. In the last two decades, a wave of early-competence findings has been publicized (e.g. Special Issue in *Cognition*, 2011), overshadowing numerous earlier findings that demonstrate a protracted development of causal competence. Such recent early-competence studies have been interpreted in support of the claim that children's causal reasoning is aided by specialized and abstract knowledge about causes and effects (e.g. Cook, Goodman & Schulz, 2011). The current findings, in contrast, question such a conclusion, adding evidence for slow-developing causal competence.

As discussed above, the obtained patterns of successful and unsuccessful performance cannot be dismissed easily on the basis of a method that is too confusing or that otherwise masks true competence. And given that multi-cue events are common in children's everyday experience (e.g. Schulz, Bonawitz & Griffiths, 2007), the current study provides an opportunity to critically evaluate the overly optimistic view on children's early causal competence. At the minimum, the current findings undermine the claim that blocking performance *per se* is indicative of sophisticated causal reasoning. If blocking performance is shown indiscriminately in multi-cue events, whether causally appropriate, inappropriate, or within a non-causal context, it implies the presence of underlying associative processes (see Hall, 1991; Kruschke & Blair, 2000; but see Ali *et al.*, 2011).

Taking it a step further, the pattern of results obtained in the current set of experiments undermines the arguments of abstract causal knowledge guiding children's causal reasoning. In contrast to some other paradigms, our task context approximated the messiness of real-life causal judgment: Children had to keep in mind the causal structure of the events, they had to sort through multiple cue–target contingencies, and they had to know when cues do and do not compete. It is precisely this kind of situation that requires causal competence. The current paradigm is therefore ideal for evaluating the relative importance of top-down abstract causal knowledge to help children sort through and distinguish between causal and non-causal relations. Finding that children were not guided by causal knowledge suggests that causal knowledge, rather than being the driving force of causal learning, might be the result of it. In fact, causal reasoning is a non-trial task, at least when children have to integrate linguistic information.

This is not to say that young children lack causal understanding altogether. After all, by 4 years of age they have been exposed to causal language and explanations on numerous occasions, and they have had ample opportunity to experience the effect of their own actions and that of others. Instead, we argue that such understanding emerges in the tight interplay with the specifics of the task context and a child's prior experiences (see also Ali, Schlottmann, Shaw, Chater & Oaksford, 2010). This same ongoing interdependent processes were shown to be implicated in numerical cognition (e.g. Mix, Huttenlocher & Levine, 1996), spatial cognition (e.g. Cheng & Newcombe, 2005; Twyman & Newcombe, 2010), category learning (e.g.

Sloutsky & Fisher, 2008; Sloutsky, Kloos & Fisher, 2007; see also Sloutsky, 2010, for an extensive review), deductive reasoning (Markovits, Fleury, Quinn & Venet, 1998; A. Morris & Sloutsky, 1998; B. Morris & Sloutsky, 2002), word learning (Samuelson & Smith, 1999; Smith, Jones, Landau, Gershkoff-Stowe & Samuelson, 2002), or shape perception (e.g. Smith, 2005). The development of causal judgment may not be different from the rest of cognitive development – a process of bottom-up learning to extract unobservable deep features and ignore salient surface features.

Acknowledgements

References

Ahn, W.-K., Gelman, S.A., Amsterlaw, J.A., Hohenstein, J., & Kalish, C.W. (2000). Causal status effect in children's categorization. *Cognition*, 76, B35–B43.

Ali, N., Chater, N., & Oaksford, M. (2011). The mental representation of causal conditional inference: causal models or mental models. *Cognition*, 119, 403–418.

Ali, N., Schlottmann, A., Shaw, C., Chater, N., & Oaksford, M. (2010). Conditionals and causal discounting in children. In M. Oaksford & N. Chater (Eds.), *Cognition and conditionals: Probability and logic in human thinking* (pp. 117–134). Oxford: Oxford University Press.

Beck, S.R., Robinson, E.J., Carroll, D.J, & Apperly, I.A. (2006). Children's thinking about counterfactuals and future hypotheticals as possibilities. *Child Development*, 77, 423–426.

Bonawitz, E.B., Ferranti, D., Saxe, R., Gopnik, A., Meltzoff, A.N., Woodward, J., & Schulz, L.E. (2010). Just do it? Investigating the gap between prediction and action in toddlers' causal inferences. *Cognition*, 115, 104–117.

Buchanan, D.W., & Sobel, D.M. (2008).Bridging the gap: children's developing inferences about object labels and insides from causality-at-a-distance. In B.C. Love, K. McRae & V.M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 64–70). Austin, TX: Cognitive Science Society.

Bullock, M. (1985). Animism in childhood thinking: a new look at an old question. *Developmental Psychology*, 21, 217–225.

Bullock, M., Gelman, R., & Baillargeon, R. (1982). The development of causal reasoning. In W.J. Friedman (Ed.), *The developmental psychology of time* (pp. 209–253). New York: Academic Press.

Cheng, K., & Newcombe, N.S. (2005). Is there a geometric module for spatial orientation? Squaring theory and evidence. *Psychonomic Bulletin & Review*, 12, 1–23.

Cheng, P.W., & Novick, L.R. (1991). Causes versus enabling conditions. *Cognition*, 40, 83–120.

Cobos, P.L., Lopez, F.J., Cano, A., Almaraz, J., & Shanks, D.R. (2002). Mechanisms of predictive and diagnosis causal induction. *Journal of Experimental Psychology: Animal Behavior Processes*, 28, 331–346.

Cook, C., Goodman, N.D., & Schulz, L.E. (2011). Where science starts: spontaneous experiments in preschoolers' exploratory play. *Cognition*, 120, 341–349.

De Houwer, J., & Beckers, T. (2002). A review of recent developments in research and theories on human contingency learning. *Quarterly Journal of Experimental Psychology B: Comparative and Physiological Psychology*, 55, 289–310.

Dickinson, A., & Burke, J. (1996). Within-compound associations mediate the retrospective revaluation of causality judgments. *Quarterly Journal of Experimental Psychology*, 49, 60–80.

Fernbach, P.M., Darlow, A., & Sloman, S.A. (2010). Neglect of alternative causes in predictive but not diagnostic reasoning. *Psychological Science*, 21, 329–336.

Fernbach, P.M., Darlow, A., & Sloman, S.A. (2011). Asymmetries in predictive and diagnostic reasoning. *Journal of Experimental Psychology: General*, 140, 168–185.

Frye, D., Zelazo, P.D., Brooks, P.J., & Samuels, M.C. (1996). Inference and action in early causal reasoning. *Developmental Psychology*, 32, 130–131.

Glymour, C. (2003). Learning, prediction and causal Bayes nets. *Trends in Cognitive Sciences*, 7, 43–48.

Gopnik, A., & Glymour, C. (2002). Causal maps and Bayes nets: a cognitive and computational account of theory-formation. In P. Carruthers, S. Stich & M. Siegal (Eds.), *The cognitive basis of science* (pp. 117–132). Cambridge: Cambridge University Press.

Gopnik, A., Glymour, C., Sobel, D.M., Schulz, L.E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: causal maps and Bayes nets. *Psychological Review*, 111, 3–32.

Gopnik, A., & Sobel, D.M. (2000). Detecting blickets: how young children use information about novel causal powers in categorization and induction. *Child Development*, 71, 1205–1222.

Gopnik, A., Sobel, D.M., Schulz, L.E., & Glymour, C. (2001). Causal learning mechanisms in very young children: two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology*, 37, 620–629.

Gopnik, A., & Tenenbaum, J.B. (2007). Bayesian networks, Bayesian learning, and cognitive development. *Developmental Science*, 10, 281–287.

Griffiths, T.L., & Tenenbaum, J.B. (2005). Structure and strength in human causal induction. *Cognitive Psychology*, 51, 334–384.

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

Hall, G. (1991). *Perceptual and associative learning*. New York: Oxford University Press.

Hoemann, H.W., & Ross, B.M. (1971). Children's understanding of probability concepts. *Child Development*, **42**, 221–236.

Kamin, L.J. (1969). Predictability, surprise, attention, and conditioning. In R.M. Church (Ed.), *Punishment and aversive behavior* (pp. 279–296). New York: Appleton-Century-Crofts.

Kruschke, J.K., & Blair, N.J. (2000). Blocking and backward blocking involve learned inattention. *Psychonomic Bulletin & Review*, **7**, 636–645.

Kushnir, T., & Gopnik, A. (2005). Children infer causal strength from probabilities and interventions. *Psychological Science*, **16**, 679–683.

Kushnir, T., & Gopnik, A. (2007). Conditional probability versus spatial contiguity in causal learning: preschoolers use new contingency evidence to overcome prior spatial assumptions. *Developmental Psychology*, **43**, 186–196.

Kuzmak, S.D., & Gelman, R. (1986). Young children's understanding of random phenomena. *Child Development*, **57**, 559–566.

Larkin, M.J.W., Aitken, M.R.F., & Dickinson, A. (1998). Retrospective revaluation of causal judgments under positive and negative contingencies. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 1331–1352.

Leslie, A.M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition*, **25**, 265–288.

Lopez, F.J., Shanks, D.R., Almaraz, J., & Fernandez, P. (1998). Effects of trial order on contingency judgements: a comparison of associative and probabilistic contrast accounts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **24**, 672–694.

Lyons, D., Young, A., & Keil, F. (2007). The hidden structure of overimitation. *Proceedings of the National Academy of Sciences, USA*, **104** (50), 19751–19756.

Markovits, H., Fleury, M.L., Quinn, S., & Venet, M. (1998). The development of conditional reasoning and the structure of semantic memory. *Child Development*, **69**, 742–755.

Mix, K.S., Huttenlocher, J., & Levine, S.C. (1996). Do preschool children recognize auditory-visual numerical correspondences? *Child Development*, **67**, 1592–1608.

Morris, A., & Sloutsky, V.M. (1998). Understanding of logical necessity: developmental antecedents and cognitive consequences. *Child Development*, **69**, 721–741.

Morris, B.J., & Sloutsky, V.M. (2002). Children's solutions of logical vs. empirical problems: what's missing and what develops. *Cognitive Development*, **16**, 907–928.

Nazzi, T., & Gopnik, A. (2000). A shift in children's use of perceptual and causal cues to categorization. *Developmental Science*, **3**, 389–396.

Nazzi, T., & Gopnik, A. (2003). Sorting and acting with objects in early childhood: an exploration of the use of causal cues. *Cognitive Development*, **18**, 299–317.

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

Pearl, J. (2000). *Causality*. Cambridge: Cambridge University Press.

Reichenbach, H. (1956). *The direction of time*. Berkeley, CA: University of California Press.

Samuelson, L.K., & Smith, L.B. (1999). Early noun vocabularies: do ontology, category structure and syntax correspond? *Cognition*, **73**, 1–33.

Schlottmann, A. (1999). Seeing it happen and knowing how it works: how children understand the relation between perceptual causality and underlying mechanism. *Developmental Psychology*, **35**, 303–317.

Schulz, L., Bonawitz, E.B., & Griffiths, T. (2007). Can being scared cause tummy aches? Naïve theories, ambiguous evidence, and preschoolers' causal inferences. *Developmental Psychology*, **43**, 1124–1139.

Schulz, L.E., & Gopnik, A. (2004). Causal learning across domains. *Developmental Psychology*, **40**, 162–176.

Schulz, L.E., Gopnik, A., & Glymour, C. (2007). Preschool children learn about causal structure from conditional interventions. *Developmental Science*, **10**, 322–332.

Shanks, D.R., Pearson, S.M., & Dickinson, A. (1989). Temporal contiguity and the judgment of causality by human subjects. *Quarterly Journal of Experimental Psychology B: Comparative and Physiological Psychology*, **41**, 139–159.

Shultz, T.R., & Altmann, E. (1982). Young children's use of transitive inference in causal chains. *British Journal of Psychology*, **73**, 235–241.

Siegler, R.S., & Liebert, R.M. (1974). Effects of contiguity, regularity, and age on children's causal inferences. *Developmental Psychology*, **10**, 574–579.

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

Sloutsky, V.M. (2010). From perceptual categories to concepts: what develops? *Cognitive Science*, **34**, 1244–1286.

Sloutsky, V.M., & Fisher, A.V (2008). Attentional learning and flexible induction: how mundane mechanisms give rise to smart behaviors. *Child Development*, **79**, 639–651.

Sloutsky, V.M., Kloos, H., & Fisher, A. (2007). When looks are everything: appearance similarity versus kind information in early induction. *Psychological Science*, **18**, 179–185.

Smith, L. (2005). Action alters shape categories. *Cognitive Science*, **29**, 665–679.

Smith, L.B., Jones, S.S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object name learning provides on-the-job training for attention. *Psychological Science*, **13**, 13–19.

Sobel, D.M., & Kirkham, N.Z. (2006). Blickets and babies: the development of causal reasoning in toddlers and infants. *Developmental Psychology*, **42**, 1103–1115.

Sobel, D.M., & Munro, S.A. (2009). Domain generality and specificity in children's causal inference about ambiguous data. *Developmental Psychology*, **45**, 511–524.

Sobel, D.M., & Munro, S.A. (2006). When Mr Blicket wants it, children are Bayesian. Proceedings of the 2006 annual meeting of the Cognitive Science Society, Vancouver, BC.

Sobel, D.M., Tenenbaum, J.B., & Gopnik, A. (2004). Children's causal inferences from indirect evidence: backwards

blocking and Bayesian reasoning in preschoolers. *Cognitive Science*, **28**, 303–333.

Sobel, D.M., Yoachim, C.M., Gopnik, A., Meltzoff, A.N., & Blumenthal, E.J. (2007). The blicket within: preschoolers' inferences about insides and causes. *Journal of Cognition and Development*, **8**, 159–182.

Téglás, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J.B., & Bonatti, L.L. (2011). Pure reasoning in 12-month-old infants as probabilistic inference. *Science*, **332**, 1054–1059.

Tenenbaum, J.B., & Griffiths, T.L. (2003). Theory-based causal inference. In S. Becker, S. Thrun & K. Obermayer (Eds.), *Advances in neural information processing systems 15* (pp. 35–42). Cambridge, MA: MIT Press.

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

Twyman, A.D., & Newcombe, N.S. (2010). Five reasons to doubt the existence of a geometric module. *Cognitive Science*, **34**, 1315–1356.

Van Hamme, L.L., & Wasserman, E.A. (1994). Cue competition in causality judgments: the role of non-presentation of compound stimulus elements. *Learning and Motivation*, **25**, 127–151.

Waldmann, M.R. (1996). Knowledge-based causal induction. In D.R. Shanks, D.L. Medin & K.J. Holyoak (Eds.), *The psychology of learning and motivation , Vol. 34: Causal learning* (pp. 47–88). San Diego, CA: Academic Press.

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

Wasserman, E.A., & Berglan, L.R. (1998). Backward blocking and recovery from overshadowing in human causal judgment: the role of within-compound associations. *Quarterly Journal of Experimental Psychology*, **51B**, 121–138.

Williams, D.A., Sagness, K.E., & McPhee, J.E. (1994). Configural and elemental strategies in predictive learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 694–709.

Zelazo, P.D., Reznick, J.S., & Piñon, D.E. (1995). Response control and the execution of verbal rules. *Developmental Psychology*, **31**, 508–517.