Integrating top-down and bottom-up approaches to children’s causal inference

David M. Sobel

Department of Cognitive and Linguistic Sciences

Brown University


Acknowledgments: I was supported by NSF (DLS-0518161 to DMS) during the writing of this chapter. I would like to thank all of the parents and children who participated in the research described here. I’d also like to thank Scott Johnson, David Buchanan, Claire Cook, Tom Griffiths, Natasha Kirkham and the members of the NYBUG workshop for helpful discussions about material in this chapter. Address correspondence to D. Sobel, Department of Cognitive and Linguistic Sciences, Box 1978, Brown University, Providence, RI, 02912. Phone: 401-863-3038. Fax: 401-863-2255. Email: David_Sobel_1@brown.edu
Integrating top-down and bottom-up approaches to children’s causal inference

Several chapters in this volume are dedicated to describing how children learn conceptual structure from the data available to them (e.g., Kirkham, this volume; Rakison & Cicchino, this volume; Sloutsky, this volume). My plan for this chapter is to focus this discussion on a particular piece of conceptual knowledge: understanding the causal relations among events. Piaget, on whom the constructivist approach to cognitive development is based, recognized the importance of causality early in his career (Piaget, 1929, 1930). However, he failed to attribute significant causal reasoning abilities to young children, with preoperational children often receiving the label “precausal” based on their verbal explanations of behaviors in the world. The first goal of this chapter is to highlight young children’s sophisticated causal reasoning abilities.

But also critical to the present discussion in this volume, there have been two approaches to causal learning. There is a long tradition of research in causal learning and inference that has focused on how causal knowledge is acquired from observing events – algorithms that construct a mental model of causal knowledge from patterns of correlational information. One might consider such a tradition more “bottom-up”. But there is also a tradition in cognitive development – dating back to Piaget – of children using their prior knowledge or contextual information in the environment to learn new causal knowledge. One might consider this tradition more “top-down.” The second goal of this chapter is to consider how to integrate these two approaches for describing children’s causal reasoning abilities.

In particular, following a set of proposals developed with Josh Tenenbaum and Tom Griffiths (Griffiths, 2005; Griffiths & Tenenbaum, 2005, 2007; Sobel, Tenenbaum,
& Gopnik, 2004; Tenenbaum & Griffiths, 2001, 2003), I will suggest a description of children’s causal inference. This approach offers a way of considering how causal principles are acquired from data to learn representations of causal structure. I will first present some background information, then some empirical work consistent with this description as well as what might be developing. Finally, I will consider some limitations of this mechanism, focusing on other information that might be available to the child to facilitate their causal inference and learning.

Causal Learning from “Bottom-up” Mechanisms

There have been numerous accounts of causal learning in which a representation of causal structure is built from observing data in the environment. On these accounts, children bring to the learning environment only the ability to translate information about associations among events into a causal representation. The simplest such account is that children associate causes and effects in the same way that animals associate conditioned and unconditioned stimuli in classical conditioning (e.g., Mackintosh, 1975; Rescorla & Wagner, 1972).

But since associative models only output strength relations, they do not appear to make predictions about how learners use causal knowledge to generate interventions to elicit effects. It appears that even rats are capable of causal reasoning in a manner that reflects more than just an associative mechanism (Blaisdel et al., 2006). As a result, several independent research programs have suggested that to generalize an associative approach, causal learning occurs by transforming a measure of associative strength into a measure of causal strength. Measures of causal strength are then used to make inferences or generate interventions. Some of these models were based on the Rescorla-Wagner
equation (see e.g., Cramer, Weiss, Williams et al., 2002). Other accounts emerged as researchers discovered a set of learning paradigms that this model has trouble explaining (e.g., Krushke & Blair, 2000; Van Hamme & Wasserman, 1994; Wasserman & Berglan, 1998). One advantage of these accounts is that they allow a way to describe how a learner might generate *interventions* on the world – actions (usually intentional) that change the value of an event exogenously (without affecting other variables in the model directly). To use a traditional example, some associative mechanisms were designed to describe classical conditional paradigms, in which the learner passively observed the environment. These accounts of human learning also take operant paradigms into account, in which the learner also generates actions, which have varying degrees of efficacy, and must learn the strength of the existing causal relation (see e.g., Dickinson & Shanks, 1995, for a detailed discussion).

Still other endeavors have taken to consider more complex relations among events beyond stimulus, response, and reinforcement (e.g., Allan, 1980; Cheng, 1997; Shanks, 1995). These models estimate the strength of a particular causal model using the probability that an effect occurs given a cause and some background information. The critical difference between these models and the ones mentioned above is that they estimate the strength of a fixed representation of causal structure, and do so accurately only given sufficiently large quantities of data (see Tenenbaum & Griffiths, 2001, for further discussion of this issue). Most of these models, however, are agnostic as to how that causal structure is fixed, with a potential exception being the power PC model (Cheng, 1997; Novick & Cheng, 2004), which suggests ways of discerning cause from effect (see e.g., Cheng & Novick, 1990).
Causal Structure Learning

The models described above focus on deriving the strength of a set of known causal relations. Informally, here is the first way in which traditionally bottom-up accounts of causal learning can be integrated with prior knowledge: If the learner is determining the strength of a known cause and effect, then there must be some knowledge in addition to the data that identifies cause from effect. While there are several theories of causal inference that place such mechanistic information central to understanding causality (Ahn et al., 1995; Shultz, 1982), such knowledge might also be entirely minimal, perhaps limited to only priority, contiguity, and contingency (e.g., Hume, 1978/1739; Michotte, 1962).

However, there are some contemporary accounts of causal learning that consider how causal structure is learned: how do adults (and children) recognize that an event is a cause or effect of another event (in addition to considering the strength of that causal relation)? Most of the psychological investigation on this approach has concentrated on adult causal learning (Griffiths & Tenenbaum, 2005; Lagnado & Sloman, 2004, Steyvers, Tenenbaum, Wagenmakers, & Blum, 2003; Tenenbaum & Griffiths, 2001; Waldmann & Hagmayer, 2005; see Lagnado et al., 2007, for a review). There is also a line of psychological investigation that suggests children construct an abstract representation of the causal structure among a set of variables (Gopnik, Sobel, Schulz, & Glymour, 2001; Schulz & Gopnik, 2004; Sobel et al., 2004; see Gopnik, Glymour, Sobel, Schulz, Kushnir, & Danks, 2004, for a review).

This description of causal learning and inference has been grounded in the literature on causal graphical models, which have been developed in computer science
Integrating top-down…and statistics (Pearl, 2000; Spirtes, Glymour, & Scheines, 2001). Causal graphical models are representations of a joint probability distribution: the probability that each possible combination of events occurs. These representations embody conditional probability information among events. Events are represented as nodes, and causal relations are represented as edges between nodes.

Making inferences from this account relies on a set of assumptions. One assumption is that any vertex represents a causal relation between two nodes, specifically in the form of a mechanism that can be either observed or unobserved (Pearl, 2000). As such, any graph is consistent with a set of probabilistic models that specify the nature of the relation among the variables. A unique description of that causal structure is made by parameterizing the graph: defining the probability distribution for each variable conditioned on its parents. Parameterizing a graph can be thought of as assigning weights to each edge. These weights represent the strength of the corresponding causal relation. A graph’s parameterization can reflect the nature of the mechanism(s) by which causes produce effects.

Causal graphical models support reasoning about interventions – actions that change the value of variables in the graph (without directly influencing those other variables, see Pearl, 2000). Consider the very simple graph: \( X \rightarrow Y \). In this graph, the probability that event \( Y \) takes a particular value given that event \( X \) takes a particular value is the same when you observe that \( X \) has that value as when you act to make \( X \) have that value. Such interventions are represented by Pearl (2000) and others as the \( do(X) \) operator. Note that the opposite is not true in this graph: The probability that \( X \) has a particular value given that you observe \( Y \) has a particular value is not necessarily the
same as the probability that X has that value given that you make Y take on the same value. To use a classic philosophical example, if I make the rooster to crow at 2am, I shouldn’t expect the sun to rise; the causal relation between sunrises and roosters crowing runs in the opposite direction (see Woodward, 2003, for further discussion). Children clearly learn causal structure from observing (and generating) these interventions (see e.g., Schulz, Gopnik & Glymour, 2007).

A second assumption that underlies causal graphical models is the faithfulness assumption. Faithfulness specifies that data are indicative of the causal structure in the world. Suppose that three events are related in the following manner: X → Y ← Z and that X has a generative relationship with Y (i.e., X occurring raises the probability that Y will occur) and that Z has a preventative relationship with Y (i.e., Z occurring lowers the probability that Y will occur). Faithfulness states that the causal relations among X, Y, and Z will never be such that X and Z exactly cancel each other’s effects on Y, so that the three events appear independent. I do not know of a psychological investigation dedicated to faithfulness; however, most psychologists investigating children’s causal learning assume this to be true.

A third assumption is the Markov Assumption, which is a way of translating between causal relations and conditional probability information (Pearl, 2000). The Markov assumption states that the value of an event (i.e., a node in the graph) is independent of all other events except its children (i.e., its direct effects) conditional on its parents (i.e., its direct causes). For example, consider the causal model A → B → C. In this model, the values of events A and C are dependent. The Markov assumption states that these values become independent conditional on the value of event B. C has no
Integrating top-down…

children, and B is its only parent. If you want to predict the value of C and know the value of B, additional knowledge about the value of A doesn’t help: the only influence that A has on C is through B.

In the next section, I will consider evidence that suggests children engage in causal reasoning in a manner consistent with the Markov assumption. Specifically, this evidence suggests that young children can recognize dependencies among events as well as when events are independent based on the presence of a third event. Such inference is tantamount to recognizing the difference between correlations due to causal relations and correlations due to spurious associations.

*Learning Causal Structure using Structure Learning: Data from Young Children*

In order to investigate whether children recognize the difference between dependence and conditional independence information, we need a method that presents a novel causal property to children wherein researchers can control the amount of prior knowledge they possess. Much of the research I will describe uses a *blicket detector* (shown in Figure 1), a machine that lights up and plays music (controlled by the experimenter) when certain objects are placed upon it. The blicket detector presents a novel, non-obvious causal property, which any object might possess.

[Gopnik et al. (2001) trained 3- and 4-year-olds that objects that activated the detector were labeled, “blickets”. Children quickly learned this relation. Then, children observed a set of trials in which objects either independently activated the machine, or did so only in the presence of another object. Specifically, on the *one cause* trials, children were shown two objects. Children observed one object (A) activate the detector...]

[Insert Figure 1 Approximately Here]
by itself. Then, they saw that the other object (B) did not activate the detector by itself. Finally, they saw objects A and B activate the detector twice together. Children were asked whether each object was a blicket. Three- and 4-year-olds labeled only object A as a blicket (although this was more likely for the older children), recognizing that object B only activated the detector in the presence of the object A.

Performance on these trials were compared with performance on two cause trials, in which the same children were shown two objects that activated the detector individually with the same frequency as the objects in the one cause trials. Specifically, children saw two new objects (C and D). Object C was placed on the machine three times and activated it all three times. Object D was placed on the machine three times, and activated two out of three times. Children categorized both objects as blickets. Both objects individually activated the detector; they just did so with different frequencies.

These data suggest that children recognize the difference between two events that are dependent because of a causal relation and two events that are dependent because of the presence of a third (causal) event. This procedure generalizes beyond reasoning about physical events: Schulz and Gopnik (2004) demonstrated that 3- and 4-year-olds make similar inferences across a variety of domains. Younger children also appear to make similar inferences. Using slight manipulations to the procedure, Gopnik et al. (2001) also demonstrated that 30-month-olds made these inferences, and Sobel and Kirkham (2006) demonstrated that children as young as 19-months generated interventions on the machine consistent with this inferential ability.

The trouble with simply concluding that children reason according to the Markov assumption is that the data presented above are analogous to blocking, a phenomenon
from the animal conditioning literature (Kamin, 1969). In a blocking procedure, a learner is shown an association between a conditioned and unconditioned stimulus (e.g., that a tone predicts the occurrence of food). This association is trained until asymptote, and then the learner is shown a novel stimulus, which presented in compound with the established conditioned stimulus will predict the same unconditioned stimulus (e.g., that the same tone paired with a light will predict food). In most cases, learners do not learn that the second stimulus is predictive. Various models of associative reasoning (e.g., Rescorla & Wagner, 1972) were designed to explain this phenomenon. In the blicket detector paradigm described above, one might consider object A to be analogous to the first stimulus, object B the second stimulus, and the detector’s activation the unconditioned stimulus. Children’s performance, thus, is analogous to that of animal learners.

It is necessary to consider alternate procedures that models of associative reasoning have difficulty explaining. One such example involves considering how children reason retrospectively about ambiguous events (following Shanks, 1985). Sobel et al. (2004) introduced 3- and 4-year-olds to the blicket machine in the same manner as Gopnik et al. (2001), and then presented them with two types of trials. In their one-cause trials, children saw that two objects (A and B) activated the machine together, and then that object A did not activate the machine by itself. In their backwards blocking trials, children saw two new objects (C and D) activate the machine together, and then that one of those objects (C) did activate the machine by itself (note that this procedure is analogous to the reverse of Kamin’s blocking procedure described above, hence its name). Children were asked whether each of these objects were blickets.
The blicket status of objects A and C are unambiguous given these data, but one’s intuitions about objects B and D should differ. Object B should be a blicket in the one-cause trial; this is consistent with the Markov assumption: object A only activates the machine dependent on the presence of object B, so object B must be the causal factor. Object D’s status is uncertain; the data (under a few assumptions, which I will make clear in subsequent sections) are equally consistent with it being a blicket and it not being one. However, these intuitions differ from the associative relations that objects B and D have with the machine’s activation. In both cases, children observe the object activate the machine in conjunction with another object. That other object (A or C) then activates or fails to activate the machine, but this piece of information should not change the associative relation between objects B and D and the machine’s activation. If children were responding on the basis of these associative calculations, they should treat these objects the same. Three- and 4-year-olds responded in a manner consistent with the intuitions, not the associative relations: Object B was almost always judged to be a blicket, while object D was judged to be so approximately 35% of the time. Preschoolers reasoned in a manner consistent with the Markov assumption, and less consistent with at least some models of causal reasoning based on calculations of associative strength (e.g., Rescorla & Wagner, 1972).

An Aside: Data from Infants

An open question raised by the previous section is whether younger children would reason in a similar manner. A variety of researchers have suggested that children’s causal reasoning abilities develop during the preschool years (e.g., Bullock, Gelman, & Baillargeon, 1982; Das Gupta & Bryant, 1989; Goswami & Brown, 1990;
Gottfried & Gelman, 2005). Further, preschoolers’ ability to relate causal inferences to perceptions of time develops after the preschool years (e.g., McCormack & Hoerl, 2005). Although Gopnik et al. (2001) demonstrated that 30-month-olds engaged in similar inferences using modified procedures, a re-analysis of their data suggested that 3- and 4-year-olds potentially extracted different information about the causal mechanisms that underlie how the blicket detector functions (Griffiths, Sobel, Tenenbaum & Gopnik, in preparation).

These findings are mostly concerned with how children understand causal mechanisms – how events are related to each other in a causal manner. In subsequent sections, I will outline a description of a domain-general causal learning mechanism, and then demonstrate how domain-specific knowledge influences children’s use of this mechanism. But before we consider those ideas, the issue of whether younger children’s causal reasoning is consistent with the Markov assumption is still open.

Sobel and Kirkham (2006) considered this question by investigating 19- and 24-month-olds’ inferences using a similar procedure to the one-cause and backwards blocking trials described above. In their one-cause condition, they placed two objects (A and B) on the machine together, which activated, and then showed the children that object A failed to activate the machine by itself. The objects and detector were slid over to the child, who was asked to “make it go.” In their backwards blocking condition, they placed three objects on the table (C, D, and E). Objects C and D made the machine go together, and then object C made the machine go by itself. Object C was removed, and the child was given objects D and E with the detector to “make it go.” Twenty-four-month-olds used object B to activate the machine in the one-cause trial, more often than
they used object D to do so in the backwards blocking trial. However, the younger children responded at chance levels, and did not discriminate between objects B and D.

A difficulty with asking children this young to make manual responses consistent with their causal knowledge (i.e., put objects on the detector to make it go), is that there are cases in which 18-month-olds fail to engage in simple imitative “means-ends” behaviors (e.g., Uzgiris & Hunt, 1975; see also Gopnik & Meltzoff, 1992). In the one-cause trial, children had to inhibit an event they observed activate the machine (placing both objects on it) in favor of a novel intervention (placing only object B on it). The demand characteristics of this experiment might have overwhelmed the toddlers from producing the appropriate inferences.

There is reason to believe that 18-month-olds, and even younger infants, have the ability to detect conditional probabilities among events. Saffran, Aslin and colleagues found that 8-month-old infants could parse a stream of auditory stimuli based solely on the transitional probabilities within and between syllables (i.e., the likelihood that one syllable would predict the next syllable; Aslin, Saffran, & Newport, 1998; Saffran, Aslin, & Newport, 1996). Infants’ statistical learning abilities extend beyond learning word boundaries. Infants are capable of recognizing and discriminating between complex grammars relating words together (e.g., Gomez & Gerken, 1999). They also parse auditory streams based on statistical probabilities even when the stimuli are tones (Saffran, Johnson, Aslin, & Newport, 1999). This suggests that the ability to perceive statistical structure is perhaps not language specific.

Further evidence for this position comes from work on infants’ parsing of visuo-spatial sequences of events. Using a procedure analogous to Saffran et al. (1996),
Kirkham, Slemmer, and Johnson (2002) demonstrated that infants as young as 2-month-old register statistical relations among sequences of visual events (see Kirkham, this volume, for a detailed description of this literature). Similarly, Fiser and Aslin (2002) demonstrated that 9-month-olds can recognize conditional probability relations between the spatial positions of visual events. Both of these projects have their origins with work by Haith and colleagues (Haith, 1993; Wentworth, Haith, & Hood, 2002), who demonstrated that young infants (3-4-month-olds) learn simple, two-location spatial sequences of events. Haith and colleagues used a visual expectation paradigm, in which infants’ eye gaze to a particular spatial location represented where the infants thought an event would appear.

Sobel and Kirkham (2006) modified this technique to investigate whether infants reasoned about sequences of events in a manner similar to preschoolers’ causal inference. Eight-month-olds were shown a video screen similar to what is seen in Figure 2. Four frames were always present on the screen. I will refer to the top center frame as A and the bottom center frame as B, the right frame as C and the left frame as D, but this was counterbalanced in the experiment. Infants observed sequences of events appear in their respective frames. During familiarization, infants saw a sequence of three types of events (the first event was randomly chosen from the three). One was that the A and B events could appear in their respective frames together. These were two grayscale events that silently rotated in space for 8 seconds. The AB compound predicted the occurrence of the C event (which occurred in the C frame) with 100% certainty. The C event was a more interesting color event, which moved around in the frame and was accompanied by a piece of cartoonish music. It also lasted for 8 seconds. The C event, however, did not
predict any event. Half of the time it was followed by the AB compound (in which case, the subsequent event would again be C), and half of the time it was followed by the D event, which was the same color event occurring in the D frame for 8 seconds accompanied by the same music. The D event also did not predict anything. Half of the time it was followed by the AB compound, and half of the time it was followed by C.

[Insert Figure 2 Approximately Here]

Infants observed this sequence of events until they saw four occurrences of the AB→C pairing (usually 11 events total). Immediately after the last AB→C pairing, infants who had been assigned to the indirect screening-off condition observed only the A event appear on the screen by itself, followed by the D event. This sequence was shown twice. Infants in the backwards blocking condition observe only the A event followed by the C event (twice). Immediately after this, infants were shown the B event, and then the screen went blank. The music that had accompanied the C and D events began to play, and at this point infants’ eye gaze was measured for 8 seconds.

One might think of this sequence of events in a manner similar to objects being placed on a blicket detector. Events A and B correspond to the two objects being placed on the detector, and events C and D correspond to the detector activating or not, respectively. In the indirect screening-off condition (analogous to the one-cause condition in the previous experiments described above), events A and B together predict C, and then A alone does not. Events A and C are dependent in the presence of B, but independent conditioned on the absence of B; this makes B predictive of C. The expectation is that infants will look more to the C frame, expecting an event to appear there. In the backwards blocking condition, the associative relation between events B
and C is the same as in the indirect screening-off condition, but since A and C are not independent conditioned on the absence of B, B is not necessarily predictive of C. We would expect to find an interaction between the amount of time spent looking at each frame and the condition the infant was assigned to. This was exactly what was found (see Figure 3). Eight-month-olds spent more time looking in the C frame in the indirect screening-off condition than the D frame, and more time looking in the D frame than the C frame in the backwards blocking condition. Moreover, they spent more time looking in the C frame in the indirect screening-off condition than backwards blocking condition.

[Insert Figure 3 Approximately here]

These data suggest that infants’ statistical learning abilities appear to be consistent with the Markov assumption. However, this inferential ability may not be available to younger infants. In a follow-up, Sobel and Kirkham (2007) found that 5-month-olds’ pattern of looking time in response to the same procedure was quite different (see Figure 3). Five-month-olds looked longer at the C frame in the backwards blocking condition, and equally long at the two frames in the indirect screening-off condition, inconsistent with the Markov assumption. One interpretation of these data is that infants might be developing a mechanism for causal and statistical reasoning that moves from recognizing associations among events to one that incorporates the Markov assumption. However, the 5-month-olds’ responses were inconsistent with an associative mechanism as well. An alternative interpretation is that when events A and B occur together, younger infants might simply treat them as the same event. If infants are treating event A alone, event B alone, and the AB compound as the same event, then their pattern of performance is consistent with both associative reasoning mechanisms and reasoning mechanisms...
consistent with the Markov assumption. More research is necessary to discriminate between these possibilities.

Second, the research on infancy presented so far has focused on infants’ statistical reasoning, and not necessarily their understanding of cause and effect. These data do not demonstrate that 8-month-olds register that event B causes event C in the indirect screening-off condition. Rather, they suggest that infants’ statistical reasoning is consistent with the Markov assumption, and may form the building block for a representation of causal knowledge. An open question is to consider how to convert this procedure to one in which infants’ causal reasoning can be measured.

Natasha Kirkham and I have begun several investigations focused on this question. One involves infants watching videos of objects placed on a blicket detector, consistent with the data shown in the one-cause and backwards blocking conditions described above. Using a violation of expectation procedure, we should be able to discern infants’ expectations about the causal efficacy of individual objects. Further, using the anticipatory eye gaze paradigm, we are attempting to train children that looks to particular locations of a screen actually cause events to occur. Using eye gaze to allow infants to generate interventions might allow them to respond in a causal manner to sequences similar to the one used in our previous work. These investigations are currently underway.

*Bayesian Inference as a Description of Causal Structure Learning*

An objection that one might have to the lines of research described above is that the difference between responses in the indirect screening-off or one cause condition and responses in the backwards blocking condition is problematic for certain models of
associative reasoning (e.g., Rescorla & Wagner, 1972), but not others. Several contemporary accounts of associative reasoning were designed with the backwards blocking paradigm in mind (e.g., Kruschke & Blair, 2000; Van Hamme & Wasserman, 1994; Wasserman & Berglan, 1998). For instance, Wasserman and Berglan (1998) use a derivative of the Rescorla-Wagner equation, in which the strength of a relation changes positively when a potential cause and effect occur and negatively when the effect occurs without a potential cause. Similarly, models of causal reasoning that rely on the estimation of causal parameters based on the frequency with which events co-occur also explain the backwards blocking data (e.g., Cheng, 1997; Shanks, 1995). These models categorize events as causes or effects and then calculate the probability that an effect occurs given a cause and some background information. For example, Cheng’s (1997) power PC model makes a clear prediction about the causal efficacy of the objects in the one-cause conditions, but generates an undefined value in the backwards blocking case, which can be interpreted as consistent with the present findings.

Is there a method of distinguishing among all of these competing options as explanations of children’s causal reasoning? One difficulty with considering the majority of these algorithms is that they rely on multiple pieces of data (i.e., large sample sizes) in order to make rational inferences. What we have observed in children’s causal reasoning is that they appear capable of making such inferences based on small amounts of data. Following researchers in adult cognition and cognitive science (e.g., Griffiths & Tenenbaum, 2005; Steyvers et al., 2003; Tenenbaum & Griffiths, 2001, 2003), Sobel et al. (2004) proposed that children’s causal learning and inference was better described by a model that relies on Bayesian inference.
On this view, causal reasoning can best be described by inference over a set of hypotheses \((H)\). Hypotheses take the form of a causal graphical model with a particular parameterization. Each hypothesis \((h_1, h_2...h_n)\) is assigned a prior probability, \(p(h)\) before observing any data. These priors reflect the learner’s causal knowledge about possible causal structures as well as any other information the learner gleams from the environment before observing the data. Given the data, \(d\), (values for the variables in the hypotheses) the learner computes the posterior probability that each hypothesis is the actual causal structure of the system, \(p(h \mid d)\). This is done using Bayes’ rule:

\[
p(h \mid d) = \frac{p(d \mid h) p(h)}{\sum_{h' \in H} p(d \mid h') p(h')} \tag{1}
\]

The prior \(p(h)\) is the probability that each hypothesis is the hypothesis that actually generated the data. The value \(p(d \mid h)\) is the likelihood of the observed data being generated if that particular hypothesis was the actual causal structure in the world. For example, if \(A \rightarrow B\) with a deterministic parameterization (i.e., \(A\) always causes \(B\)) is one of the hypotheses, and the data consists of trials of \(A\) occurring in the absence of \(B\), then the \(p(d \mid h) = 0\) for this particular hypothesis. This hypothesis requires \(B\) to occur whenever \(A\) occurs, and that is not the case.

To see this computational description in action, consider the backwards blocking sequences in which two objects activate the blicket detector together and then one of those two objects activates the detector by itself. There are four hypotheses potentially consistent with these data:

- \(h_1\): that neither object is a blicket
- \(h_2\): that only the first is a blicket
$h_3$) that only the second is a blicket

$h_4$) that both are blickets

The data are equally inconsistent with hypotheses $h_1$ and $h_3$ (i.e., $p(d \mid h) = 0$), since the first object has to be a blicket (it activates the machine by itself, but more on this in the subsequent sections). The data, however, are equally consistent with the other two hypotheses ($h_2$ and $h_4$), and as such the $p(d \mid h) = 1$ for both. But this description allows for another piece of information to influence causal inference, namely the prior probabilities, and these priors might affect children’s inferences.

A rational way in which these priors might be assigned is through observing the base rate of objects with causal efficacy – the frequency of blickets in the world. If there are few blickets in the world, then the prior probability of hypothesis $h_2$ should be higher than that of $h_4$, since $h_2$ posits fewer blickets. Similarly, if blickets are relatively common, then the reverse should be true. Using this logic, Sobel et al. (2004) presented 3- and 4-year-olds with a version of the backwards blocking procedure in which they initially manipulated the base rate of blickets. Children were shown the blicket detector, and taught that blickets make the machine go. The experimenter then brought out a box of identical blocks. In one condition (the rare condition), 2 out of the first 12 blocks shown to the child activated the detector, and were categorized as blickets. In the other condition (the common condition), 10 out of the first 12 blocks activated the detector, and were blickets. Then the experimenter brought out two more blocks (A and B), and proceeded with the backwards blocking demonstration: these blocks together activated the detector, and then that object A activated the detector by itself.
The causal status of object A is unambiguous – it is a blicket – and all of the children categorized it as such. The causal status of object B is ambiguous given the data, but if children relied on the observed base rates, they should treat this object differently between the rare and common conditions. In the common condition, both 3- and 4-year-olds claimed that object B was a blicket, consistent with children recognizing prior probabilities when evaluating ambiguous data. In the rare condition, the 4-year-olds claimed that object B was not a blicket, again consistent with recognizing priors, but the 3-year-olds did not. They judged that the B object was a blicket regardless of the base rate of blickets in the environment.

There are two conclusions from these data. The first is that 4-year-olds’ inferences were consistent with the Bayesian description in so far as they could recognize priors from the environment and use that information to make rational inferences about ambiguous data. The second is that there was a developmental difference between 3- and 4-year-olds’ inferences. It is possible that a system for causal inference develops between these ages. However, there is another possibility, which involves considering what information is necessary for the child to possess in order to formulate a hypothesis space accurately.

Children’s Developing Knowledge about Blicket Detectors

In the previous section, I asserted that there were four hypotheses consistent with the backwards blocking data. What knowledge was necessary to form this hypothesis space? Do children possess this knowledge?

Some spatiotemporal knowledge appears necessary. First, placing an object on the blicket detector makes it activate; the detector’s activation shouldn’t cause the
interner to place an object on it. Second, an object’s location in space should be
independent of another object’s locations in space. Given research on infants’ causal
perception (e.g., Leslie & Keeble, 1987; Oakes & Cohen, 1990), and preschoolers’ causal
knowledge (e.g., Bullock et al., 1982; Sophian & Huber, 1984), it seems reasonable to
assume that young children reason according to these two principles. Such knowledge
limits the hypothesis space to the four models described above.

But children also need to understand that there is a particular parameterization
between objects and the detector activating – what Tenenbaum and Griffiths (2003) and
Sobel et al. (2004) called the activation law. Do children recognize that there is
something about a blicket that makes the machine go? The activation law specifies that
children recognize that there is some mechanism that relates blickets to the detector’s
activation in a deterministic (or near deterministic) manner. This information allows the
learner to recognize that the data in the backwards blocking procedure are ambiguous.
Without this information (i.e., if children believed that blickets only sometimes made the
machine go), the data are more consistent with object B having the capacity to activate
the detector than not. To illustrate this, suppose that blickets only activated the detector
80% of the time. Even though object A clearly is a blicket (by virtue of activating the
machine alone), it might have failed to be responsible for activating the machine when it
and object B were placed on the machine together; there would be a nontrivial chance
that the detector’s activation was uniquely caused by object B having the efficacy to
activate the machine.

While there is some good evidence that suggests 4-year-olds treat causal relations,
including causal relations involving the blicket detector, as deterministic (Bullock et al.,
1982; Schulz & Sommerville, 2006), it is not clear whether younger children do so as well. Further, even this work does not suggest that children recognize that deterministic data are related to particular causal mechanisms. As such, in a series of investigations, my colleagues and I considered how 3- and 4-year-olds reasoned about the relation between the causal properties of artifacts and non-obvious, internal properties. Our question was whether 3- and 4-year-olds recognized that insides of objects could act as mechanisms for those objects’ causal properties, and whether children might understand such mechanisms differently across domains of knowledge.

In one set of experiments (Sobel, Yoachim, Gopnik, Meltzoff, & Blumenthal, 2007), we presented children with the blicket detector (although we simply labeled it as a machine, so that children were not influenced by object label information) and a set of objects such as those shown in Figure 4. Two objects were externally identical, and another was unique in appearance. All three objects had holes drilled into them, covered by dowels, which could reveal whether each contained an internal part. Four-year-olds were shown the insides of each toy: one of the identical objects and the unique object contained an internal part (a white map pin), while the third member of the set was empty inside. Children were then shown that the member of the pair with the internal part activated the detector, and they were asked to show the experimenter another object that would activate the machine. The majority of children chose the other object with the internal part (66% of the time), significantly more often than chance. Four-year-olds also claimed that objects that shared internal parts were more likely to share causal properties (i.e., activate the detector) than objects that shared external parts (e.g., had stickers on them).
This inference also worked in the other direction. In another experiment 3- and 4-year-olds were shown the same sets of objects and the blicket detector (again, without it being labeled as such), and were shown the causal efficacy of the three objects. One member of the pair and the unique object activated the detector, while the other member of the pair did not. The member of the pair that activated the detector was opened to reveal that it contained an internal part, and children were asked which other object also contained such an inside. A striking developmental difference was found: 3-year-olds chose the other object that activated the detector 31% of the time, significantly lower than what would be expected by chance. Four-year-olds chose this object with significantly greater frequency (72%), and more often than chance responding. Importantly, Sobel et al. (2007) also ran a condition in which the association between the detector’s activation and each object was held constant, but the object was not causally related to the machine. Each object was held over the detector, and the experimenter pressed a button on the detector for the objects that would have activated the machine. Here, both 3- and 4-year-olds made causal responses less than 30% of the time, significantly lower than what would be expected by chance.

What these data suggest is that 4-year-olds, but not 3-year-olds, recognize that there is a relation between an object’s causal and internal properties. However, these data do not demonstrate that 4-year-olds understand an activation law – that there is something about the internal part that is responsible for activating the detector. Four-year-olds do integrate some amount of correlational information together when making inferences about causal mechanisms: they only respond on the basis of the machine’s
activation when the spatial-temporal connection between the object and machine warrant a causal relation. A stronger argument would be to demonstrate that 4-year-olds, but not younger children, interpret an object’s internal parts as being necessary and sufficient for the detector’s activation.

To test this, Emily Blumenthal and I introduced 3- and 4-year-olds to the blicket detector and provided children with (what we thought was) the strongest possible information about its efficacy. We told children that the machine was a “blicket machine” and “things with blickets inside made the machine go”. We then showed children that a set of objects with internal parts (labeled blickets) all activated the machine, and that a set of objects without internal parts did not. We also showed children that the detector activated if at least one object with a blicket inside was on it. After receiving this training, children were shown two objects (A and B), which activated the machine together. The door on object A was opened to reveal it was empty. Children had no trouble inferring whether each object contained an internal part. The critical question was an intervention question – children were asked to make the machine go. The child has observed the experimenter generate an intervention that activates the machine – placing both objects on it. But this imitative response is not the most efficient way of activating the machine. If children recognize that the internal part is responsible for the object’s causal property, then they should recognize there is no need to put object A on the detector, and when asked to generate an intervention that activates the machine, place only object B on the machine. This was the response generated by the majority of 4-year-olds, significantly more often than the younger children. The younger children were more likely to imitate, and place A and B on the machine together.
These data suggest that 4-year-olds tie together the correlational information they observe between each object and the detector’s activation, and mechanism information about what is necessary for each object to activate the detector: namely, a non-obvious property. Three-year-olds have a harder time integrating this information. This development might relate to children’s developing use of a Bayesian mechanism of causal inference. Correlational information is reflected in how the data generate posterior probabilities of each hypothesis being correct. Mechanism information is reflected in how those hypotheses are formed: what causal structures come under consideration and how those causal structures are parameterized.

Specifically, what we would like to show is that children who recognize the activation law, by virtue of connecting objects’ causal and internal parts together, are more likely to engage in inferences consistent with recognizing the prior probability information they observe. The 3-year-olds who failed to discriminate between the rare and common conditions in Sobel et al.’s (2004) procedure might have lacked the understanding that the way an object can produce its causal properties can be related to its insides. Lacking this knowledge might indicate that they lacked an activation law relating objects with the detector, which would make their failure to respond like the older children rational.

That’s Mr. Blicket to you: Causal Mechanisms across Domains

In order to test this hypothesis, we need to consider how we might facilitate 3-year-olds’ understanding of an activation law. One possibility is to consider how children reason about such causal relations in another domain of knowledge; all of the experiments mentioned so far have been exclusive to the domain of blicket detectors and
aspects of physical causality. Can similar manipulations be performed in another
domain? There is some reason to suspect that some causal inference abilities, such as
inferences consistent with the Markov assumption, are domain-general (e.g., Schulz &
Gopnik, 2004). However, these inferences involved general logical principles, not
specific pieces of causal mechanism information.

Research in theory of mind tells us that young children understand a particular
mental state – the results of an agent’s desires – at very young ages. Eighteen-month-
olds recognize that others can have desires different from their own (Repacholi &
Gopnik, 1997). Two- and 3-year-olds have a good understanding of the outcomes of
fulfilled and unfulfilled desires (Wellman & Woolley, 1990). Three-year-olds can also
keep track of their own and other’s desires over time and changes in the environment
(e.g., Gopnik & Slaughter, 1991). Fawcett and Markson (2005) asked 2-year-olds to
make inferences about their own preferences based on another’s desires. They showed
children that one person consistently played with toys that matched the child’s
preferences, and that another person consistently played with toys that did not match the
child’s preferences. They then presented two novel (equally preferential) toys, and each
person played with one. When those two toys were given to the children, they preferred
to play with the toy associated with the first person. This suggests that children infer a
non-obvious property to the toy based on another’s desires, since that property must be
responsible for those desires (see also Perner, 1991; Yuill, 1984).

Sobel and Munro (2006) manipulated the blicket detector to attempt to introduce
it to 3-year-olds as a psychological agent. They placed a set of cardboard eyes on the
machine (shown in Figure 5) and introduced it to children as “Mr. Bicket.” The
Integrating top-down…

The experimenter conducted a dialogue with the machine, which activated spontaneously in response to questions and comments (this procedure was modeled after Johnson, Slaughter, & Carey, 1998; Johnson, Booth, & O’Hearn, 2001, who used a similar procedure to study agent gaze following in infants). The children were then told that they were going to play a game in which Mr. Blicket would tell them whether he liked an object. They then repeated the procedure used by Sobel et al. (2007) to study whether 3-year-olds linked the internal parts of objects with their causal property (in this case, whether Mr. Blicket liked the object). Three-year-olds did link the causal property with the object’s insides in this condition, significantly more often than in another condition, in which the same procedure was performed with a machine that spontaneously activated during the warm-up of the procedure, with the same temporal contiguity as Mr. Blicket’s activation (70% vs. 41% of the time).

These data suggest that 3-year-olds might integrate the correlational data they observe about agent’s desires towards objects with mechanism information – that there must be something about those particular objects responsible for Mr. Blicket’s desires. If this is the case, then 3-year-olds might have an activation law about Mr. Blicket’s desire, and reason more consistently with the Bayesian description than they do in the physical domain. Research in the lab right now is considering this question (Sobel & Munro, in preparation). We introduced 3-year-olds to Mr. Blicket in the same manner as described above, and then gave them the same rare or common training as in Sobel et al. (2004), followed by the same backwards blocking trial. Two new objects (A and B) were placed on Mr. Blicket together, and he activated. Then object A alone was placed on him with
the same result. All children claimed that Mr. Blicket liked object A, and the question was how they categorized object B.

Three-year-olds claimed that Mr. Blicket liked object B 44% of the time when trained that he liked relatively few things (recall the base rate in this condition was 1/6). In contrast, when Mr. Blicket liked many things (a base rate of 5/6), children responded that he liked object B 93% of the time, a significant difference between the conditions. Performance in the rare condition, however, could have been influenced by a number of factors. One possibility is that children were influenced by the spontaneous activation of the box, and would respond in a similar manner to a blicket machine that they observed spontaneously activate. This was not the case. Another group of 3-year-olds were shown a blicket machine that spontaneously activated during the initial part of the procedure. They were trained that objects that activate the machine (and hence, are blickets) were rare, and were given the same procedure. In this condition, 3-year-olds categorized object B as a blicket 72% of the time, more often than in the desire condition.

Similarly, another possibility is that children were simply more interested in Mr. Blicket than they were a blicket machine. There are cases where children’s interest level clearly mediates their cognition (e.g., Renninger & Wozniak, 1985). To consider this possibility, we gave another group of 3-year-olds the same Mr. Blicket procedure, except that we labeled his activation as indicating what he was thinking about, instead of what he liked. Unlike desire, 3-year-olds have little knowledge of other’s belief states (e.g., Wellman, Cross, & Watson, 2001), the role of thinking in other mental activities (Flavell et al., 1995; Johnson & Wellman, 1992; Lillard, 1996), or the possibility that thoughts could be related to objects or other thoughts (Eisbach, 2004). It seemed likely that only a
few 3-year-olds would recognize that an agent’s thoughts could be based on an internal property of objects, which would provide them with a causal mechanism equivalent to the activation law. This also appeared to be the case. In this condition, 3-year-olds categorized object B as something Mr. Blicket was thinking about 72% of the time, more often than in the desire condition.

Further, in the three conditions in which children were trained that the causal power (Mr. Blicket’s desire, thoughts, or the activation of a spontaneous machine) was rare, we also gave children a set of unrelated cognitive measures as well as a measure in which they were asked to relate the causal power of objects to those objects’ insides (analogous to the procedure used in Sobel et al., 2007). Across all of these conditions, the ability to relate the causal property of objects to those objects’ insides predicted whether children claimed that object B did not have causal efficacy (i.e., a response consistent with the Bayesian description), even when age and other measures of general cognition were considered.

These data indicate that children are not specifically developing causal inference abilities between the ages of 3 and 4. Rather, children appear to have such an inferential mechanism in place at the age of three, and lack the particular domain-specific knowledge necessary to use that mechanism appropriately. The Bayesian description I am suggesting here (following similar proposals by Griffiths & Tenenbaum, 2005; Tenenbaum & Griffiths, 2001, 2003; Tenenbaum et al., 2007) offers a rational way of considering how children’s developing prior knowledge influences their causal reasoning abilities.
An open question is how such causal knowledge might be acquired. In the final sections, I want to consider two possibilities. The first is an extension of the Bayesian mechanism. The second attempts to integrate other pieces information from the environment.

*Learning Causal Mechanisms*

So far we have considered how children recover a representation of the causal environment based on the data they observe. This learning mechanism is guided by a particular set of causal principles, which potentially constrain the hypothesis space children consider and the parameterization of those hypotheses. An open question is how children develop knowledge of these causal principles.

Consider the mechanism that underlies the blicket detector. The previous sections argued that preschoolers develop a conception that the mechanism that underlies the detector’s activation is deterministic. This knowledge is what allows us (and young children) to make inferences based on small samples of data. In almost all of the experiments described above, children are never shown data that contradicts a deterministic mechanism. What happens if this is the case?

In Gopnik et al. (2001), children were shown cases in which objects sometimes made the machine go and sometimes did not. In their *two-cause* trials, children inferred that an object that activated the blicket detector two out of three times was a blicket most of the time. This trial provides evidence that the detector is not deterministic, and might activate based on a more probabilistic mechanism. How might seeing this trial first affect children’s inferences on other trials, in which a deterministic mechanism is required?
Like the backwards blocking procedure, Gopnik et al’s (2001) one-cause procedure relies on children understanding that there is a deterministic mechanism that relates blickets to the blicket detector (recall that on a one-cause trial, object A activates the machine by itself, object B does not by itself, than then both objects activated the machine together twice). If the detector is probabilistic, then there should be the possibility that object B is a blicket; object B might have failed to be effective when it was placed on the machine alone, but demonstrated its efficacy when placed on the machine with object A. If the detector is deterministic, then this is not the case: object A should be a blicket, and object B should not be by virtue of it failing to activate the detector independently. Gopnik et al. (2001) found that overall, children (particularly 4-year-olds) who were shown these data responded consistently with the deterministic interpretation. Tom Griffiths and I reanalyzed performance on the one-cause trials as a function of whether they observed a two-cause trial first (recall in the two-cause trial, one object activates the detector probabilistically; it fails to activate the machine the first time it is placed on it, and does so the next two times). Four-year-olds were more likely to say that object B was a blicket in the one-cause trial if they saw a two-cause trial first.

Griffiths et al. (in preparation) considered more systematically whether children and adults can extract mechanism information from the data they observe. Specifically, if learners first observe evidence that the detector is deterministic will they make different inferences about the same data than if they first observe evidence that the detector is not deterministic? This question can also be formulated as one of Bayesian inference, although the hypotheses are about the principles that govern how hypotheses about causal models are formulated. In this example, the hypotheses include the nature
of the activation law – the mechanism that relates objects to the detector – in addition to the specific causal structures (following Tenenbaum, Griffiths, & Kemp, 2006; Tenenbaum et al., 2007). For purposes of space, I will only describe the psychological investigation with young children, but we have done similar investigations on adults.

Griffiths et al. (in preparation) showed 4-year-olds the blicket detector, and trained them that the detector was either deterministic or probabilistic. In the deterministic condition, children were introduced to the detector as in Gopnik et al. (2001). They then observed six objects, each placed on the machine three times. Five of the six objects activated the machine all three times, and were labeled blickets; the other object failed to activate the machine all three times, and was labeled as not a blicket. In the probabilistic condition, children received the same introduction, and saw the same six objects. But here, the objects that activated the machine perfectly in the previous condition did so with some noise. Objects either activated the machine perfectly (one object), two out of three times (two objects) or one out of three times (two objects), and any of the objects that activated the machine was labeled a blicket. The object that failed to activate the machine all three times was still labeled as not a blicket, keeping the base rate of blickets the same across the conditions.

Children then observed a set of trials similar to the one cause condition in Gopnik et al. (2001). The critical part of the trial involved them observing two new objects (A and B). Object A activated the machine by itself once. Object B failed to activate the machine by itself once, and then A and B together activated the machine twice. Children were asked whether each was a blicket. In the deterministic condition, performance paralleled Gopnik et al (2001): children stated that object A was a blicket (100% of the
time), and object B was not (only 9% of the time). In the probabilistic condition, children stated that object A was a blicket (92% of the time), but were significantly more likely to state that object B was as well (79% of the time).

These data offer preliminary evidence that 4-year-olds not only can recover information about causal models from the data that they observed, but that they also recover the principles necessary to learn causal structure from those data. Given the same correlational information, their inferences were different dependent on the nature of the mechanism they were exposed to. Children’s understanding of these mechanisms might not be terribly deep; they might not have explicit understanding of the mechanism, but rather just be aware that some kind of mechanism exists, which constrains inference certain ways. This seems consistent with the work on relating causes and insides: the internal parts of the objects in Sobel et al (2007) are dummy mechanisms, but the older children treat them as if they were the mechanism for the detector’s activation, without (apparently) a real conception of how such mechanisms function.

**Integrating Top-Down and Bottom-Up Learning**

Appealing to a Bayesian description of a causal learning mechanism – specifically one that might be able to extract such mechanisms knowledge from observed data – does not imply that all causal learning is “bottom up.” Instead, the Bayesian description seems more integrative: “top-down” principles for constraining causal learning can be derived from data, but this should not be considered the only way causal learning works. Below, I suggest several additional ways children might be able to acquire information about the principles for causal learning.
Testimony. More likely than not, the best manner in which children learn new causal structures (or new causal principles) is through direct instruction – what Harris and Koenig (2006) call learning from “testimony.” Harris, Pasquini, Duke, Asscher, and Pons (2006), for example, demonstrated that children made strong ontological commitments about different non-observable scientific and endorsed entities (e.g., vitamins vs. Santa Claus). Further, the degree of their commitment in these entities varied with the exposure that they received about them.

More generally, one could imagine that children learn a great deal of causal structure simply by being told about that structure (something that might be particularly important in learning science, see Klahr & Nigam, 2004). This is evident in the introduction to most blicket detector experiments, in which children are told that machine is a “blicket machine,” and that objects that make it go are “blickets.” The fact that children learn this readily (established in the pretests of almost all of these experiments), suggests that they can learn causal principles directly from the language they hear, but this is a topic for further investigation.

Analogy. Numerous investigations suggest that young children can make inferences from analogies (e.g., Brown & Kane, 1988; Gentner, 1977), and this is especially true when reasoning about causal relations (e.g., Goswami & Brown, 1989; Goswami, Leevers, Pressley, & Wheelwright, 1998; Ratterman & Gentner, 1998). This suggests that children can come to make new causal inferences from analogous information, or learn new information faster/more accurately if the analogy is mapped out for them. Emily Hopkins and I (Hopkins & Sobel, 2007; Sobel & Hopkins, submitted) have recently considered this possibility by looking at a particular type of causal
inference: recognizing the difference between a generative cause and an enabling condition. Specifically, we found that 4-year-olds struggled to understand enabling conditions in a decontextualized environment (where the part of an object that acted as the enabling condition was labeled an “inside”). However, young children do appear to understand enabling conditions in a particular setting: a CHILDES (MacWhinney, 2000) analysis revealed that children talk about how batteries are necessary to make machines and toys function. Four-year-olds were able to make proper inferences about enabling conditions in a condition in which the part that acted in this manner was labeled as a battery. Further, we found that such reasoning generalized across domains: presenting information in the form of a battery affected 4-year-olds reasoning about enabling conditions in the psychological domain as well and their inferences about physical events.

*Contextual Information in Data.* A limitation of the causal graphical model framework is that it does not easily describe a way in which contextual cues can influence learning. For example, active construction of knowledge in the world is a hallmark of both classic (e.g., Montessori, 1912; Piaget, 1952) and certain contemporary (e.g., Gopnik & Meltzoff, 1997) approaches to cognitive development. The computational approaches described here do not consider whether the child has an active hand in constructing their knowledge as opposed to recovering causal structure from simply observing the environment.

The ability to control what data one observes, and generate interventions consistent with those data appear to facilitate learning over observing similar data in adult participants (Lagnado & Sloman, 2004, Steyvers, Tenenbaum, Wagenmakers, & Blum,
Integrating top-down…

2003; Waldmann & Hagmayer, 2005), in child participants (Schulz et al., 2007), and in animals (Blaisdell et al., 2006). Moreover, young children appear to treat their own data as more informative than if the same data were generated by another person (Kushnir & Gopnik, 2005).

Jessica Sommerville and I investigated how young children’s causal learning was affected by particular contextual demands (Sobel & Sommerville, in preparation). We found that 4-year-olds whose free play with a system allowed them to discover causal structure learned that structure better than children whose free play with a system came after they observed an experimenter generate a small number of interventions on the system (but enough that discovered the structure). Further, when children were shown identical intervention data, which was sufficient to learn a causal structure, children who were given an inappropriate rationale for why the experimenter was generated those data failed to learn the system; children given an appropriate rationale learned above chance values (Sobel & Sommerville, submitted). These contextual factors are not part of the computational description I’ve described so far, and must be accounted for therein.

Conclusions

In this chapter, I have suggested a description of causal inference based on Bayesian inference, which illustrates how children engage in causal learning (for a more detailed description of this model, see Griffiths & Tenenbaum, 2007). This description is meant at the computational level of analysis (followed Marr, 1982), which means that an obvious limitation of this approach is that it should not be taken for the actual algorithm by which children learn causal knowledge, nor should it be considered how the brain
integrates such inference. However, in describing the way in which children learn causal knowledge, we provide insight into these questions.

I want to conclude by emphasizing that computational models are a good way to focus an investigation, but a psychological description of human causal learning should not be completely model-dependent (whether that model be bottom-up, top-down, or something in between). One should integrate model with human workings to describe psychological accounts of reasoning (what Lagnado et al., 2007, calls a “heuristic-based” approach). Here, it should be emphasized that young children possess considerable causal reasoning abilities, starting at a very young age. The goal of future research is to describe these abilities, and potentially an algorithmic and implementational level of children’s causal inference – in more detail.
References


Integrating top-down…


Integrating top-down…


Figure Captions

Figure 1. A blicket detector (specifically the detector used in Gopnik & Sobel, 2000, and elsewhere). In this case, an object is placed on the detector, and it is enabled, so that the object is activating the detector. This particular detector lights up red and plays *Für Elise*.

Figure 2. A screenshot shown to infants in Sobel and Kirkham (2006, 2007). In this shot, events A and B are presented together.

Figure 3. Amount of time spent looking to the C and D frames in the indirect screening-off and backwards blocking conditions by 8-month-olds (Sobel & Kirkham, 2006) and 5-month-olds (Sobel & Kirkham, 2007).

Figure 4. Stimulus set used to measure whether children appreciated the relation between objects causal properties and insides (Sobel et al., 2007).

Figure 5. Mr. Bicket
Figure 1
Figure 2
Figure 3

Mean Looking Times to the Frames

5-month-olds
(Sobel & Kirkham, 2007)

8-month-olds
(Sobel & Kirkham, 2006)

Condition

Backwards Blocking
Indirect Screening-Off
Backwards Blocking
Indirect Screening-Off

C
D

time (ms)
Figure 4
Figure 5