

9

Interactions Between Causal and Statistical Learning

David M. Sobel & Natasha Z. Kirkham

Our everyday interactions with the world suggest that there is much causal knowledge that children must discover. Causal knowledge enables children to interpret the current state of the world rationally and to engage in predictive inference and explanation. Traditionally, young children's causal knowledge has been considered "perceptually driven" or "precausal" (e.g., Piaget, 1929). Contemporary research, however, has shown that young children's causal reasoning abilities are actually quite sophisticated. Infants recognize causal properties of objects, including containment, support, and contact (e.g., Hespos & Baillargeon, 2001; Leslie & Keeble, 1987; Needham & Baillargeon, 1993; Spelke, Breinlinger, Macomber, & Jacobson, 1992). Before their second birthday, toddlers recognize various nonobvious causal relations, especially about others' desires and intentions (e.g., Meltzoff, Gopnik, & Repacholi, 1999). By age 5, children understand that biological and psychological events rely on nonobvious, hidden causal relations (e.g., Gelman & Wellman, 1991; Gopnik & Wellman, 1994). More generally,

preschoolers recognize the importance of Hume's principles—temporal priority, spatial priority, and contingency—in making judgments about causal relations (Bullock, Gelman, & Baillargeon, 1982; Shultz, 1982). Preschoolers also appear to have sophisticated explanatory and counterfactual reasoning abilities (Harris, German, & Mills, 1996; Schult & Wellman, 1997; Sobel, 2004; Wellman & Liu, chapter 16, this volume).

As developmentalists, we wish to describe how children learn causal knowledge and develop their reasoning abilities. How children represent and acquire causal knowledge, however, is an interdisciplinary question, and this volume illustrates how philosophy, computer science, and cognitive psychology can offer different insights into the process. We would like to suggest that other branches of developmental research—specifically research on infants' statistical learning—offer insight into causal learning. Conversely, understanding how children learn and reason about causal relations might provide insight into other areas of development.

vedq1\

In this chapter, we examine the relationship between young children's causal learning and inference abilities and their capacity to perceive the statistical associations between salient events. Of course, there are significant differences between recognizing statistical relations and causal knowledge. Knowing causal relationships allows learners to generate explanations and reason about counterfactuals, neither of which is supported by pure statistical association. Statistical associations do allow for simple predictions about future events and the chunking of correlated stimuli (for more efficient processing). But, causal knowledge goes much further than that, allowing for a "calculus of intervention" (Pearl, 2000): inferences about the outcome of intentional manipulative actions that change the state of events in the world. Recognizing that two events are associated provides no information about the result of interventions on either event. But, although correlations do not equate to causal relations, they are often a good place to start. Understanding whether and how children acquire statistical information about events in the world should provide a starting point for researchers interested in causal learning. Likewise, children's causal reasoning abilities might provide insight into phenomena discussed in the statistical learning literature.

How Statistical Regularity Can Translate to Causal Knowledge

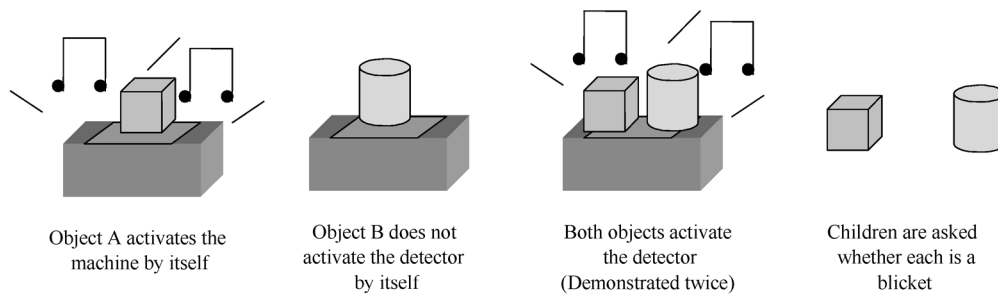
A system for causal learning has a particular problem: Although some causal relations seem directly perceivable—such as watching a ball launch another ball (Michotte, 1963)—in general, causal knowledge is not directly perceptible. One goal of research in children's causal learning has been to describe how causal knowledge can be recovered from the environment. How children recognize correlations between objects and events seems a good place to start. For example, knowing that a particular causal relation exists suggests that certain data will occur; if event *X* causes event *Y*, then the occurrence of *X* will make the occurrence of *Y* more likely (all other things being equal). Observing such correlations might offer insight into causal structure. Seeing that *Y* is more likely in the presence of *X* than in its absence often leads us to conclude that *X* is a cause of *Y*. Indeed, some adult experiments on causal learning suggest that such probabilistic reasoning might be considered a normative model of causal inference (Allan, 1980; Shanks, 1995).

Of course, such correlations do not always equate to genuine causal conclusions. Consider three events related by a simple causal chain $X \rightarrow Y \rightarrow Z$. In this situation, *X* and *Y* are correlated, *X* and *Z* are correlated, and *Y* and *Z* are correlated. Temporal priority (or other forms of prior knowledge) might inform you of the directions of the potential causal relations specified by these correlations, but the correlations themselves potentially overgeneralize the causal structure. Whether *X* causes *Z* directly or only indirectly through *Y* is ambiguous given only this information. What is necessary is a system that recognizes not only the dependencies among these events, but also their conditional independencies as well. Observing that *X* and *Z* only co-occur in the presence of *Y* (and thus are independent in the absence of *Y*) suggests the causal chain model. If this conditional independence relationship was absent, then a more general model in which *X* directly causes both *Y* and *Z* (and in which *Y* causes *Z*) is more likely.

What this suggests is that children must recognize the dependencies among events as well as conditional probability information to learn the causal structure of the world. Researchers in causal learning have examined whether children recognize conditional probability information when making causal inferences (Gopnik, Sobel, Schulz, & Glymour, 2001; Sobel, Tenenbaum, & Gopnik, 2004). Much of this research introduced children to a *blicket detector*, a machine that lights up and plays music when certain objects are placed on it. The blicket detector presents a novel, nonobvious property of each object: its activation potential. (The machine is actually controlled through an "enabling" switch. When the switch is on, any object will activate the detector. When it is off, no object will activate the detector.)

Gopnik et al. (2001) trained 3- and 4-year-olds that objects that activated the detector were called *blickets*. Children quickly learned this association. Then, children observed a set of trials in which objects either independently activated the machine or did so only dependent on the presence of another object. Specifically, in the one-cause trials, children were shown two objects. Children observed one object *A* activate the detector 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 (see Figure 9-1 for a schematic of this procedure). Children were asked whether each object was a blicket. In this condition, 3- and 4-year-olds

One-Cause Condition



Two-Cause Condition

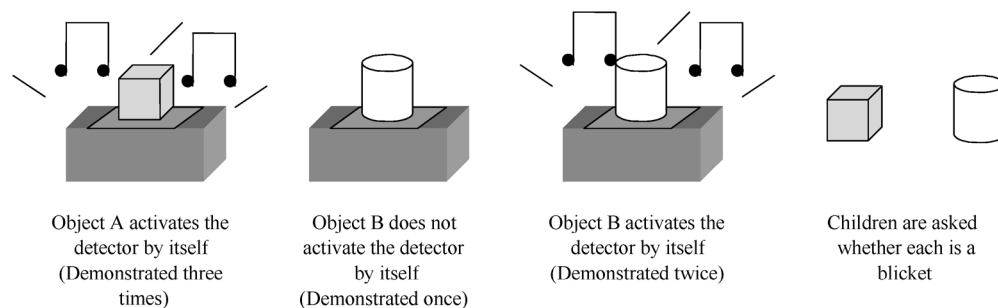


FIGURE 9-1 Schematic representation of the procedures used by Gopnik et al. (2001, Experiment 1).

labeled only object A as a blicket. Object B only activated the detector in the presence of the object A.

Performance on these trials was compared with performance on the two-cause trials, in which children were shown two objects that both independently activated the detector with the same frequency. 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 it two of three times. On these trials, 3- and 4-year-olds categorized both objects as blickets. Both objects independently activated the detector; they just did so with different frequencies.

These data suggest that children can recognize the difference between dependencies and conditional independencies between two events when faced with information about their statistical regularity (what Reichenbach in 1956 called *screening-off* reasoning). This type of reasoning represents a move from recognizing just the co-occurrence among events to recognizing the information necessary to make causal inferences. This procedure generalizes beyond

reasoning about physical events: Schulz and Gopnik (2004) demonstrated that 3- and 4-year-olds make similar screening-off inferences across a variety of domains (see also Schulz et al., chapter 5, this volume). Younger children also appear to make similar inferences. Using slight manipulations to the procedure, Gopnik et al. (2001) demonstrated that 30-month-olds made these screening-off inferences. Sobel and Kirkham (2005) demonstrated that children as young as 19 months also reasoned in this manner about objects placed on a blicket detector.

The Associative Challenge

The trouble with the procedure described is that a mechanism for causal reasoning does not exclusively explain children's ability to make screening-off inferences. Screening off is a form of *blocking*, a phenomenon from the animal conditioning literature. In a classic blocking experiment (Kamin, 1969), an animal is shown an association between a conditioned and an unconditioned stimulus (e.g., that a

tone predicts the occurrence of food). This association is trained until asymptote, and then the animal is shown a novel conditioned stimulus presented in compound with the established stimulus (e.g., that the same tone paired with a light will predict food). Animals do not learn that the light is predictive. One interpretation of these data is that the animals recognize that light only predicts food in the presence of an established predictor (i.e., the tone). Various models of associative reasoning (e.g., Rescorla & Wagner, 1972) were designed to explain this phenomenon.

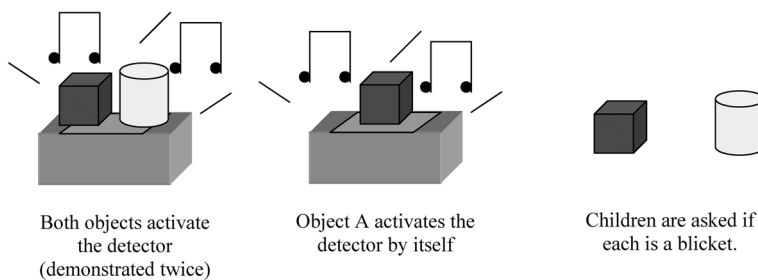
However, inferential models that rely on calculating the associative strength among events have difficulty when the data involve learners making retrospective inferences. One example, taken from the contingency judgment literature, is the phenomenon of *backward blocking* (Shanks, 1985; Shanks & Dickinson, 1987). In these experiments, adult learners were presented with two stimuli in compound (A and B) that elicited some effect. Learners were then shown that one of those two events alone (A) elicits the effect. Given this information, adults rated that the B stimulus did not have the causal efficacy necessary to produce the effect.

Can children engage in backward blocking about causal events? Sobel et al. (2004) introduced preschoolers to the blicket detector and trained them that

blickets activated the machine. They showed children two objects (A and B) that activated the machine together. Then, they showed children that Object A activated the machine by itself. This procedure is shown in Figure 9-2. The critical question was how children would rate Object B. Its causal status is uncertain. If children engage in backward blocking, then they should determine that it is not a blicket. This was the case: Children rarely labeled Object B as a blicket.

Sobel et al. (2004) also demonstrated that children did not follow a simple algorithm that only recognized associations among events (e.g., Rescorla & Wagner, 1972). In a different type of trial—indirect screening-off trials—children were shown two different objects (C and D) that activated the machine together and then that Object C failed to activate the detector. This procedure is also shown in Figure 9-2. In this circumstance, only Object D should be considered a blicket: Object C fails to activate the detector independently, so the only logical conclusion children could draw is that Object D has the causal efficacy necessary to activate the detector. Indeed, 3- and 4-year-olds generated this response. However, responding on the basis of only associations, one would consider Objects B and D's associative strength with the detector's activation to be the same. In both

Backwards Blocking Condition



Indirect Screening-off Condition

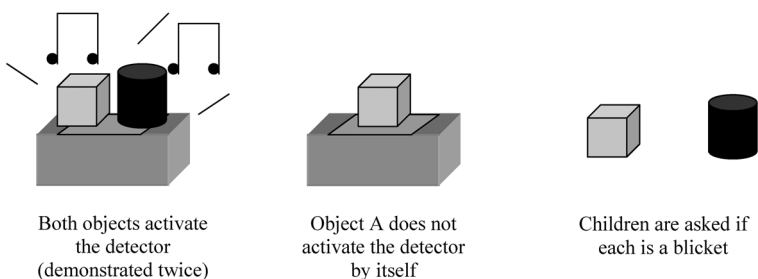


FIGURE 9-2 Schematic representation of the procedures used by Sobel et al. (2004, Experiment 1).

cases, the object activates the detector with another object. The efficacy of that other object alone should have no bearing on its associations with the detector.

Although these data are inconsistent with various associative accounts, there are several different categories of learning algorithms that describe how adults make inferences about contingencies among events, which can explain these findings. Children might recognize causal relations based on a calculation of associative strength among events that relies on more complicated associative mechanisms (Dickinson, 2001; Kruschke & Blair, 2000; Wasserman & Berglan, 1998). These models were designed with the backward blocking phenomenon in mind. Alternatively, children might make causal inferences through estimates of causal strength based on the frequency with which events co-occur. Such models, like the ΔP model (Allan, 1980; Jenkins & Ward, 1965; Shanks, 1995) and the PowerPC model (Cheng, 1997, 2000), calculate an estimate of the strength of a presumed causal relationship given a set of data. The backward blocking data are also consistent with these possibilities.

Sobel et al. (2004) and others (Tenenbaum & Griffiths, 2003, this volume; Tenenbaum, Sobel, & Gopnik, 2005) pointed out that many of these learning mechanisms rely on multiple exposures to data (i.e., large sample sizes). Models that calculate causal structure through associative strength or through parameter estimation, like the ΔP and PowerPC models, must have large sample sizes to function properly. In this view, these backward blocking data are inconsistent with all of these accounts because children make inferences based on relatively small sample sizes. However, Sobel et al. (2004) also wanted to demonstrate that children's causal inferences were based on a different type of learning algorithm: one that relies on Bayesian inference. In Bayesian inference, learners assign a probability value to a set of potential causal hypotheses and then update the values of those probabilities given the observed data based on the application of Bayes' rule. The resulting posterior probabilities are a rational estimate of the likelihood of each hypothesis being the correct causal model (see Tenenbaum & Griffiths, this volume, for a more detailed description of this model).

This account relies on the assumption that children assign the initial probabilities of each hypothesis nonrandomly: Those priors are set by the base rate of blickets. If children recognize that there are many blickets out there in the world, then hypotheses that specify that many objects are blickets should have a

higher initial probability than hypotheses that specify few objects are blickets. In this case, the hypothesis in which both Object A and Object B are blickets should have a higher initial probability than the hypothesis that only Object A is a blicket. Both are consistent with the observed data, and thus both will be updated equally by the application of Bayes' rule. Thus, the hypothesis that both objects are blickets will have a higher posterior probability. Thus, Bayesian reasoning predicts that if children know there are many blickets in the world, then they should not demonstrate backward blocking.

To test this hypothesis, Sobel et al. (2004, Experiment 3) showed children a set of identical objects that were placed on the blicket machine. They trained children that blickets were either rare or common: 12 objects were scanned 1 at a time, and either 2 or 10 activated the machine in the rare and common conditions, respectively. Then, they presented children the same backward blocking procedure with two new objects (from the same set). When 4-year-olds were trained that blickets were rare, they demonstrated backward blocking: The uncertain object was not categorized as a blicket. When 4-year-olds were trained that blickets were common, they did not demonstrate backward blocking: The uncertain object was categorized as a blicket. There was not enough counterevidence to exclude the hypothesis that both objects were blickets.

These data were qualitatively consistent with the Bayesian account. The same ambiguous backward blocking data were presented across the conditions, and children relied on the base rates of blickets to make an inference about an object with causal powers that were uncertain. Further research demonstrated that adults also reason about such data in a similar manner (Tenenbaum et al., 2005). Tenenbaum et al. also introduced a new learning problem in which both children and adults observed only ambiguous data. Adult learners were introduced to a machine like a blicket detector (a detector that responded to a special kind of lead in pencils, dubbed *superlead*, and hence, *superpencils*) and were trained that the occurrence of pencils containing this lead was rare (using the same manipulation as that of Sobel et al. in 2004—by showing them that 2 of 12 pencils chosen at random from a set activated the detector). Then, they were shown 3 pencils taken from the set at random (A, B, and C). Objects A and B activated the machine together, and then Objects A and C activated the machine together. Participants were asked to rate the likelihood that each object was a super pencil after each event.

The Bayesian model predicts four levels of performance given these data. First, ratings of Object A at the end of the trial should be highest, but not at ceiling. This reflects the fact that learners did not unambiguously observe Object A activate the machine, but the majority of hypotheses consistent with the data suggest that Object A is a super pencil. The ratings of A and B after they are placed on the machine together should be slightly lower. This reflects the fact that the data suggest that at least one of those objects must be a superpencil. The ratings of B and C at the end of the trial should be lower, but still higher than the initial ratings of each object. There are some hypotheses consistent with B and C being superpencils (namely, the hypothesis that B and C are superpencils, and A is not). Tenenbaum et al. (2005) observed exactly this four-level pattern of responses. This pattern of performance also extends to children. In a subsequent experiment, they found that 4-year-olds made similar responses, consistent with these qualitative predictions of the Bayesian model.

In general, these experiments integrate children's (and adults') use of statistical information from the environment with their causal inferences. In the rare-common backward blocking manipulation, children's retrospective inferences were guided by the base rate of blickets. Because the blicket detector introduced a novel causal relation, children had to rely on that initial exposure to establish the training. Indeed, the original backward blocking experiments can be reanalyzed in terms of the base rate of blickets. In Sobel et al.'s first experiment, the base rate of blickets was exactly 50%. In this case, 4-year-olds rated the B object (the blocked object) as a blicket 13% of the time in the backward blocking condition. In their second experiment, the base rate of blickets was slightly higher (it varied between 60% and 80%, with a mean of 68%). In this case, 4-year-olds categorized this object as a blicket 35% of the time, slightly higher than the previous experiment. Even without explicit training, 4-year-olds seemed to pick up on the base rate of objects that activated the detector and used that information to guide their inferences.

Can Younger Children Make Retrospective Inferences?

Sobel and Kirkham (2005) examined whether toddlers were capable of retrospectively making screening-off inferences. They introduced 19- and

24-month-olds to the blicket detector and established that both age groups would place causally efficacious objects on the machine. Then, children were shown two objects (A and B) that activated the machine together, and then that object A did not activate the machine by itself. When these objects and the machine were presented to the child with the instruction to "make it go," 24-month-olds placed Object B on the detector by itself significantly more often than all other responses put together. The 19-month-olds, in contrast, responded no differently from chance.

Sobel and Kirkham (2005) also presented these children with a backward blocking inference. Of course, because the children were too young for verbal measures, they could not replicate the Sobel et al. (2004) procedure. Instead, they showed children three objects (A, B, and C). Objects A and B activated the machine together, and then Object A activated the machine by itself. Object A was removed from the display, and Objects B and C and the machine were presented to the child. If children made a backward blocking inference, then they might be more inclined to choose Object C (the novel object) in this condition because they would infer that Object B is ineffective. Both 19- and 24-month-olds chose between Objects B and C at chance.

The importance of this procedure, however, is not in these results, but in the comparison with the indirect screening-off procedure because the associative strength of Object B is the same across the two tasks (at least on many associative models like the Rescorla-Wagner model). The 24-month-olds' use of Object B to activate the detector differed between these two conditions; 19-month-olds chose Object B with the same frequency across the two trials. Importantly, Sobel and Kirkham (2005) did find that these 19-month-olds recognized screening-off inferences that involved no retrospection. The critical question is whether these causal reasoning abilities developing during the toddler years.

Statistical Learning

A difficulty with testing toddlers' causal inferences is that there are some 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). Although the children who participated in Sobel and Kirkham's (2005) experiment were

slightly older, to count as making a retrospective inference in the indirect screening-off trials, they 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 these inferences.

There is reason to believe that 18-month-olds, and even younger children, 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).

For example, Saffran et al. (1996) presented infants with a 2-minute constant speech stream of 12 unique syllables, which could be parsed into four 3-syllable "words." These words were presented through speakers located on either side of the seated infant and were defined only by the transitional probabilities between syllables (i.e., there were no pauses or other cues to word beginnings or endings). Syllables that occurred within words always predicted each other; their transitional probabilities were always equal to 1. In other words, the first syllable in a word always predicted the second syllable, and the second syllable always predicted the third. Syllables that occurred across word boundaries were less predictable. In this particular case, because there were only four words in the speech stream, the transitional probability was equal to .33. The last syllable in a word predicted the first syllable of the three other words with equal likelihood. Infants were conditioned to turn their heads toward the speaker producing the novel strings (using a preferential head-turn paradigm). When they turned away from the speaker, the speech stream would stop. In this way, infants controlled their individual exposure to the auditory stimuli.

After familiarization, infants listened to three syllables that made up words [i.e., three syllables A, B, and C, in which $p(B|A)=1$ and the $p(C|B)=1$], alternating with three syllables that did not make up a word (i.e., three syllables that did not obey these transitional probabilities). The infants showed significantly greater interest in the nonwords than in the words, as measured by the amount of time spent looking at the speakers. Because infants will consistently look longer at novel stimuli, postfamiliarization (Bornstein, 1985),

these results suggest that the infants discriminated between the words and the other stimuli based on learning the transitional probabilities defining word boundaries (see also Aslin et al., 1998, for evidence that the results stem from true computation of input statistics rather than simple frequency counting).

Infants' statistical learning abilities extend beyond learning word boundaries. Infants are capable of recognizing and discriminating between complex grammars relating words together. Using the preferential head-turn paradigm, Gomez and Gerken (1999) exposed 12-month-olds to a subset of novel strings produced by one of two artificial grammars. These grammars differed only in terms of the ordering of word pairs: Individual words in the two sets and the starting and ending words were always the same. The only cues to recognition were contained in the transitional probabilities inherent in the word order. After familiarization to the grammar, infants were exposed to novel words embedded in either the original grammar or a novel grammar. Infants showed significantly increased looking time to the speaker producing novel words in the original grammar, suggesting that they could discriminate between the two grammars even when the words were unfamiliar.

The ability to extract regularities in sequential input does not seem to be a language-specific mechanism, but exists broadly across audition. Infants parse auditory streams based on statistical probabilities even when the stimuli are tones (Saffran, Johnson, Aslin, & Newport, 1999). Further, at least one species of non-human primates, cottontop tamarins (which never develop humanlike language skills), can learn statistically structured sounds (Hauser, Newport, & Aslin, 2001). This suggests that the ability to perceive statistical structure is perhaps not language specific.

There is evidence from other paradigms that infants show some sensitivity to visual spatial relations among repetitive events. Young infants learn simple two-location, predictable spatial sequences in a visual expectation paradigm (Haith, 1993). Infants also show sensitivity to spatial contingency in temporal sequences. Wentworth, Haith, and Hood (2002) presented 3-month-old infants with a spatiotemporal sequence in which a stimulus appeared on the left, in the center, or on the right of a computer monitor. Infants viewed either a fixed or a random pattern of locations, and in some cases there was a contingent relation between the identity of the central stimulus and the location of the next peripheral picture. The fixed sequence of three locations resulted in more eye movement anticipations,

and there were more anticipatory saccades to the correct location when there was a contingent relation between central and peripheral events.

Infants can also recognize statistical structure in displays of greater complexity than simple two- and three-location events. Kirkham, Slemmer, and Johnson (2002) demonstrated that infants as young as 2 months old could learn temporal sequences of shapes that were defined by transitional probabilities. Kirkham, Slemmer, and Johnson (2004) found that 8-month-olds were capable of extracting these probabilities even when the visual sequence was both temporal and spatial. In addition, Fiser and Aslin (2003) demonstrated that 9-month-olds are capable of picking up on the correlations between individual visual elements in a series of static multielement scenes. After being exposed to a number of these scenes, the infants were shown isolated element pairs that had co-occurred either frequently within the scenes or rarely; infants were capable of discriminating between the two.

Statistical Learning Across Modalities

These data suggest that infants—perhaps as young as 2 months—recognize conditional probabilities among events and respond to sequences based on those transitions. These learning and inferential abilities go beyond observing sequences of events; knowledge about the environment requires correctly correlating events across sensory modalities.

Indeed, infants develop a variety of intersensory capacities that allow them to integrate information across modalities. Newborns bind auditory stimuli to visual stimuli and then expect that the sounds and their associated objects will move together (Morrongiello, Fenwick, & Chance, 1998; Richardson & Kirkham, 2004). By 4 months of age, infants perceive the bimodal nature of objects (Spelke, 1979, 1981), and they can perceive speech bimodally (Kuhl & Meltzoff, 1982). Four-month-olds also match faces with voices based on age, gender, and (at 5 months) affective expression of the speaker (Bahrick, Netto, & Hernandez-Reif, 1998; Walker, 1982). By 5 months, infants also recognize the importance of this sensory integration. Bahrick and Lickliter (2000) demonstrated that infants habituated to a bimodal presentation of an event sequence (e.g., a hammer tapping out a particular rhythm) would dishabituate to the unimodal presentation of that information (e.g., just the visual of the hammer tapping, without the sound).

These capacities indicate that infants not only prefer multimodal cues that present them with statistical redundancies but also recognize their importance in perceiving the world. Infants' sensitivity to cross-modal information stands in contrast to the sparse, unimodal presentations of many laboratory experiments described here (e.g., Fiser & Aslin, 2003; Kirkham et al., 2002; Saffran et al., 1996). If experimental studies do not fully exploit the cross-modal sensitivity of infants, then perhaps they risk underestimating the full capacity of their learning abilities. Bahrick, Lickliter, and colleagues have presented evidence that *intersensory redundancy*, the overlap of information provided by amodal stimuli, drives selective attention (e.g., Bahrick & Lickliter, 2000; Bahrick, Lickliter, & Flom, 2004). Can infants usefully integrate statistical information across different modalities?

One way in which this question can be answered is in considering infants' understanding of objects as enduring across space and time behind an occluder. In experimental settings, typically the demonstration is unimodal (e.g., a silent visual display of a ball traveling across the visual field and passing behind and then reemerging from an occluder). Kirkham and Johnson (2004) demonstrated that 4-month-old infants, who are right at the beginning of a transition toward success at perceptual completion in an object constancy paradigm (e.g., correctly perceiving the constant trajectory the ball), benefit greatly from the presence of cross-modal information. They incorporated a continuous moving sound into the ball-and-occluder paradigm such that the sound traveled with the object from one side of the occluder to the other side. When given these multiple, redundant, cross-modal cues, 4-month-old infants could anticipate trajectories as well as 6-month-olds in a unimodal condition.

Multiple redundant cues are useful when one has to learn from probabilistic information. For example, if you test positive for a disease on a blood test that is 90% effective on two separate occasions, then you can be more than 90% sure that you are indeed suffering from the disease (up to 99% sure, assuming independent probabilities). Research modeling language learning has shown that multiple probabilistic cues (e.g., lexical stress, phonemes, and pauses) can be integrated to produce faster learning of word boundaries and syntax, even though each cue individually might be unreliable (Christiansen, Allen, & Seidenberg, 1998; Christiansen & Dale, 2001). Further, models have shown a particularly robust effect of cross-modal information in the service of learning (de Sa & Ballard, 1998).

Kirkham, Slemmer, and Johnson (2004) demonstrated one method in which redundant cue integration benefited infants' statistical learning. When 8-month-olds were presented with a visuospatial pattern (e.g., a red circle that appeared in one of six locations and in a statistically probable pattern), they were unable to learn the statistical relationships within the sequence successfully. However, when redundant color and shape cues were added into the sequence (e.g., each shape in the pattern had a unique color and shape), performance improved significantly. Redundant information supported infants' statistical learning abilities.

How Statistical Learning Informs Our Understanding of Causal Learning

These studies provide compelling evidence that infants are sensitive to statistical regularities across various modalities but leave open the intriguing question of how such abilities could support the complex inferences that exist in causal reasoning. When preschoolers use conditional probability to make judgments about whether objects areblickets, are they relying on the same mechanisms as infants learning word boundaries or structural information about the visual world?

Several different research groups have suggested that children's and adults' causal knowledge and reasoning abilities can be modeled by a particular computational framework: causal graphical models (Glymour, 2001; Gopnik et al., 2004; Lagnado & Sloman, 2004; Waldmann & Hagmayer, 2001). The data on children's causal inferences are all consistent with this representation of causal knowledge. To make these models causal, they must meet a set of assumptions (see Gopnik et al., 2004), but at heart, causal graphical models represent joint probability distributions—the frequency with which all possible combinations of events occur. This would imply that recognizing statistical regularities among events is critical for causal learning and reasoning, and infants' statistical learning abilities build up to an understanding of causal relations among events.

One implication of this hypothesis is that infants should be able to engage in the kind of retrospective inferences about statistical regularity among events. Our previous investigations suggested that 19-month-olds could not make these kinds of inferences when presented with the blicket detector procedure. However, these difficulties could have resulted from

the motor demands of the experiment. Using statistical learning procedures that involve measuring infants' eye gaze eliminated these demands. We have begun investigating this hypothesis by presenting 8-month-old infants with a statistical learning procedure that examines these abilities (Sobel & Kirkham, 2005, Experiment 2). Our procedure is shown in Figure 9-3a to 9-3c. In both conditions, 8-month-olds observed a sequence of four events. During the familiarization stage (Figure 9-3a), two of these events (A and B) always occurred together and predicted the occurrence of another event (C) with 100% frequency. The C event equally predicted a fourth event (D) or the AB compound. Likewise, the D event was equally predictive of C or AB. A sound effect (the same one) accompanied the C and D events.

After this familiarization, which lasted until infants observed the AB→C sequence four times, infants observed that one member of that compound (B) predicted either the C or the D event (Figure 9-3b). After observing these data, infants were presented with the other member of the compound (A), followed by a blank screen (Figure 9-3c), and the sound effect that accompanied the C and D events was played. Infants' eye gaze was measured for an 8-second period. When the B event did predict the C event on its own, infants were faced with a similar backward blocking inference concerning the A event; when B did not predict C, the data were similar to the indirect screening-off procedure used with the blicket detector.

We observed a significant interaction between looking time to the C and D locations and experimental condition. When infants were presented with the backward blocking data, they looked more often to the D location than the C location. The data suggest that the infants did not believe that the A event predicted the C event, even though they observed no evidence to the contrary. When infants were presented with the indirect screening-off inference, the pattern of looking times was reversed: Infants looked longer to the C location than the D location. Critically, following the A event, infants' looking times to the C location were different between the two conditions, suggesting that they were not responding on the basis of a simple associative mechanism.

The present data are inconsistent with certain models that might underlie recognizing statistical regularities. In particular, the hypothesis that children's reasoning is based solely on recognizing associations does not seem to provide the proper framework to explain these data. Similarly, models that rely

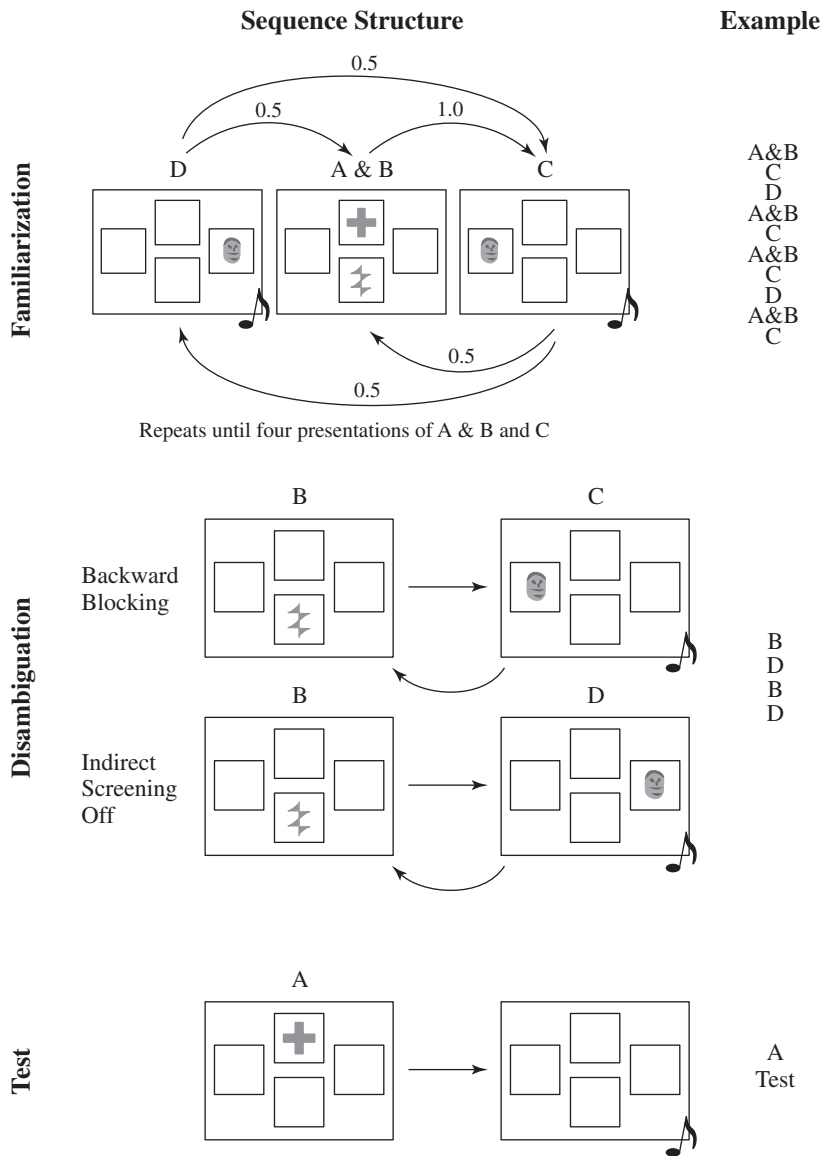


FIGURE 9-3 Schematic representation of the procedure used by Sobel and Kirkham (2005, Experiment 2). During the familiarization phase, infants were shown that Events A and B co-occurred and were always followed by Event C, but that no other event predicted any other. After four occurrences of the $AB \rightarrow C$ sequence, infants observed that the B event alone was followed by Event C (backward blocking condition) or Event D (indirect screening-off condition). After two of these sequences, infants were shown Event A, and their looking time was measured.

primarily on calculations of associative strength that do not distinguish between forms of retrospective inference, such as the Rescorla-Wagner (1972) equation and others based on it (e.g., Cramer et al., 2002), seem inconsistent with the present data.

These inferential abilities are consistent with the hypothesis that children recognize conditional probability and engage in screening-off inferences at early ages. However, unlike the blicket detector experiments, which showed that preschoolers' causal inferences

could not be explained by a variety of alternative models of causal reasoning, the present data are consistent with models of causal learning that rely on causal strength designed with retrospective inferences in mind (e.g., Kruschke & Blair, 2000; Wasserman & Berglan, 1998) as well as various parameter estimation models (e.g., Allan, 1980; Cheng, 1997; Shanks, 1995).

Like the experiments on preschoolers, there is one aspect of these data that is inconsistent with these models: Infants appear capable of making these kinds of inferences based on a small sample of data. Estimates of causal strength and measures of parameter estimation require a relatively large amount of data to make a meaningful estimation. In the present experiment, infants could do so with only four trials with the compound AB event and two trials with one of those events in isolation. However, a stronger argument would be to present infants with inferences that would be inconsistent with the models listed, parallel to the method used in previous research on preschoolers' causal inference (Sobel et al., 2004; Tenenbaum et al., 2005). We are currently attempting to determine whether one of these models best describes infants' abilities to recognize statistical regularities among events.

New Directions for Integrating Causal Learning With Statistical Learning

In addition to attempting to map out how infants recognize co-occurrences among events, we believe there are several other interactions between the causal and statistical learning literature that are worthy of future investigation. This list is not meant to be exclusive or exhaustive. Rather, we wish to articulate particular relations between the statistical learning and causal learning literature and suggest that each can benefit from discussions with the other.

The Problem of Multimodal Integration

The literature on infants' multimodal integration suggests that redundant information supports infants' statistical learning abilities. Are similar effects found in children's causal inferences? Does redundant information benefit children's understanding of causal relations?

This question has been examined indirectly by researchers interested in relation of the role of causal properties to conceptual development. Gopnik and Sobel (2000) examined whether children would extend

a novel label to objects that shared the same causal properties. They introduced children to the blicket detector without using that description. They showed children four objects and demonstrated each on the blicket detector. Critically, in "conflict" trials, two identical pairs of objects were used, and one of each activated the detector. The experimenter then labeled one of the objects that activated the detector a blicket and asked the child to show him the other blicket. The 3- and 4-year-olds chose between the perceptually identical and causally identical object with equal frequency.

Nazzi and Gopnik (2000) replicated this experiment, but added a critical piece of redundant information: They pointed out either the causal or the perceptual features of each object. When an object was placed on the detector in the causal condition, the experimenter said, "Look, it activates the detector," and in the perceptual condition, the experimenter said, "Look, this one is red." They found that children in the causal condition made more causal responses on these conflict tasks than children in the perceptual condition or those in a baseline condition.

These data suggest that children's inferences about category membership are influenced by redundant information (the machine's activation and the experimenter's language). However, there is little research investigating what information would be considered redundant and at what ages children are sensitive to this information.

The Problem of Constraining Statistical Learning

A good deal of evidence suggests that infants can recognize correlations among environmental factors and use that information to make inferences. For example, Younger and colleagues (Younger, 1990; Younger & Cohen, 1983, 1986; Younger & Gottlieb, 1988) suggested that, by the age of 10 months, infants recognize correlations among object features. A question that emerges from this discussion is whether infants are capable of detecting *any* correlation or whether constraints must be in place to guide the child.

This question has been investigated across a number of laboratories, and often developmental differences emerge. Younger children appear more capable of detecting any kind of correlation; older children only detect correlations that have some theoretical rationale (see, e.g., Madole & Cohen, 1995; Rakison, 2004). For instance, Madole and Cohen found that both 14- and

150 CAUSATION AND PROBABILITY

18-month-olds detected the co-occurrence between the form and function of an object part. However, 14-month-olds could also detect a correlation between the part of an object and the function of another part of that object. Although there are many objects in the world in which a part's form and function co-occur, the latter co-occurrence has little bearing on reality. Indeed, 18-month-olds did not detect this correlation.

There are a variety of theoretical interpretations of these data, from a top-down, theory-driven approach that suggests features correlate based on a set of explanatory principles (e.g., Murphy & Medin, 1985) to a bottom-up approach to conceptual development motivated by detecting which correlations are critical to category membership (e.g., Smith & Heise, 1992). In a discussion of these data, Madole and Oakes (1999) state that "the child's own experience acting on and observing objects is probably the primary instigator of developmental change" (p. 289). How this occurs remains an open question.

The Problem of Setting Priors

Tenenbaum and Griffiths (this volume) provide chapter and correct authors.; Tenenbaum et al., 2005) presented a Bayesian algorithm that accounts for much of the data on both preschoolers' and adults' causal inferences presented here (Sobel et al., 2004; Tenenbaum et al., 2005; see also Griffiths & Tenenbaum, in press). An important aspect of this account is that children (at least by age 4) use statistical information to set the probability of particular causal hypotheses. In these experiments, children use the frequency with which particular events occur to set their prior for each hypothesis.

However, other information about the way objects and events causally interact may be detectable from statistical information in the environment. Griffiths (2005) reexamined the original Gopnik et al. (2001) screening-off experiment, in which 3- and 4-year-olds were shown the examples of one-cause and two-cause trials shown in Figure 9-1. Children received two of each trial in a random order. Griffiths suggested that the order in which these trials were presented might have presented different information about the nature of the blicket detector. If children observe a two-cause trial first, in which one object's causal power is probabilistic, then children might interpret the detector as a probabilistic device. When they then observe the one-cause trial, in which Object B does not activate

the detector by itself once, children might interpret Object B as a blicket that simply failed on that trial (because, after all, it was shown to activate the detector with Object A two times subsequently).

Griffiths (2005) found that performance of 4-year-olds demonstrated this particular order effect. Their performance on the first one-cause trial depended on whether it was the first test trial or if they had observed a two-cause trial previously. If children observed a two-cause trial before, then they were more likely not to make a screening-off inference (i.e., to say that Object B was a blicket). Here, 4-year-olds are not recognizing the statistical regularity among events but rather that patterns of data suggest how new data could be interpreted. Younger children did not show this pattern of response. Is this developmental difference robust, or does it reflect something specific about the blicket detector paradigm? Griffiths and Sobel (in preparation) are currently investigating this question systematically. But, an open question remains: How else might the data children observe influence which causal inferences they make?

Concluding Thoughts

What infants know about the statistic of an environment has been a seminal question in language learning for some time (Aslin et al., 1998; Jusczyk & Aslin, 1995; Saffran et al., 1996). Questions about children's knowledge and use of statistical regularities in the environment have also motivated research in conceptual development (e.g., work by Younger and colleagues) as well as what infants know about object concepts (Johnson, Amso, & Slemmer, 2003; Johnson, Bremner, Slater, et al., 2004; Spelke & Van de Walle, 1993). These questions have also begun to permeate the field of theory of mind, examining what infants know about statistical regularities in detecting intentions (Baldwin, Baird, Saylor, & Clark, 2001; Brand, Baldwin, & Ashburn, 2002) or pretending (Lillard & Witherington, 2004).

What these literature bodies all have in common is that they describe some type of causal knowledge that children develop. There seem to be several places in which children's causal learning and their statistical learning abilities interact and can inform each other. Mapping out those interactions, both generally and in specific domains of knowledge, is critical to a set of exciting new research questions that can be asked.

ACKNOWLEDGMENTS We would like to thank the parents and children who have made all of the research presented here possible. We would also like to thank Alison Gopnik, Daniel Richardson, Laura Schulz, and Jessica Sommerville for comments on this chapter.

References

- Allan, L. G. (1980). A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society*, 15, 147–149.
- Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, 9, 321–324.
- Bahrick, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental Psychology*, 36, 190–201.
- Bahrick, L. E., Lickliter, R., & Flom, R. (2004). Intersensory redundancy guides the development of selective attention, perception, and cognition in infancy. *Current Directions in Psychological Science*, 13, 99–102.
- Bahrick, L. E., Netto, D., & Hernandez-Reif, M. (1998). Intermodal perception of adult and child faces and voices by infants. *Child Development*, 69, 1263–1275.
- Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, A. M. (2001). Infants parse dynamic action. *Child Development*, 72, 708–717.
- Bornstein, M. H. (1985). Habituation of attention as a measure of visual information processing in human infants: Summary, systematization, and synthesis. In G. Gottlieb & N. A. Krasnegor (Eds.), *Measurement of audition and vision in the first year of postnatal life: A methodological overview* (pp. 253–300). Norwood, NJ: Ablex.
- Brand, R. J., Baldwin, D. A., & Ashburn, L. J. (2002). Evidence for “motionese”: Modifications in mothers’ infant-directed action. *Developmental Science*, 5, 72–83.
- Bullock, M., Gelman, R., & Baillargeon, R. (1982). The development of causal reasoning. In W. J. Friedman (Ed.), *The developmental psychology of time* (pp. 209–254). New York: Academic Press.
- Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, 104, 367–405.
- Cheng, P. W. (2000). Causality in the mind: Estimating contextual and conjunctive power. In F. Keil & R. A. Wilson (Eds.), *Explanation and cognition* (pp. 227–253). Cambridge, MA: MIT Press.
- Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes*, 13, 221–268.
- Christiansen, M. H., & Dale, R. A. C. (2001). Integrating distributional, prosodic and phonological information in a connectionist model of language acquisition. In *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.
- Cramer, R. E., Weiss, R. F., Williams, R., Reid, S., Nieri, L., & Manning-Ryan, B. (2002). Human agency and associative learning: Pavlovian principles govern social process in causal relationship detection. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology*, 55B, 241–266.
- de Sa, V. R., & Ballard, D. H. (1998). Category learning through multimodality sensing. *Neural Computation*, 10, 1097–1117.
- Dickinson, A. (2001). Causal learning: Association versus computation. *Current Directions in Psychological Science*, 10, 127–132.
- Fiser, J., & Aslin, R. N. (2003). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 15822–15826.
- Gelman, S. A., & Wellman, H. M. (1991). Insides and essence: Early understandings of the non-obvious. *Cognition*, 38, 213–244.
- Glymour, C. (2001). *The mind’s arrow*. Cambridge, MA: MIT Press.
- Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, 70, 109–135.
- 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, 1–30.
- Gopnik, A., & Meltzoff, A. N. (1992). Categorization and naming: Basic-level sorting in 18-month-olds and its relation to language. *Child Development*, 63, 1091–1103.
- Gopnik, A., & Sobel, D. M. (2000). Detectingblickets: How young children use information about causal properties in categorization and induction. *Child Development*, 71, 1205–1222.
- Gopnik, A., Sobel, D. M., Schulz, L., & Glymour, C. (2001). Causal learning mechanisms in very young children: 2-, 3-, and 4-year-olds infer causal relations from patterns of variation and co-variation. *Developmental Psychology*, 37, 620–629.
- Gopnik, A., & Wellman, H. M. (1994). The theory theory. In L. Hirschfield & S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture*

- (pp. 257–293). New York: Cambridge University Press.
- Griffiths, T. L. (2005). *Causes, coincidences, and theories*. Unpublished doctoral dissertation, Stanford University, Stanford, CA.
- Griffiths, T. L., & Tenenbaum, J. B. (in press) Elemental causal induction. *Cognitive Psychology*.\vedq10\
- Haith, M. M. (1993). Future-oriented processes in infancy: The case of visual expectations. In C. Granrud (Ed.), *Visual perception and cognition in infancy* (pp. 235–264). Hillsdale, NJ: Erlbaum.
- Harris, P. L., German, T., & Mills, P. (1996). Children's use of counterfactual thinking in causal reasoning. *Cognition*, 61, 233–259.
- Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition*, 78, B53–B64.
- Hespos, S. J., & Baillargeon, R. (2001). Infants' knowledge about occlusion and containment events: A surprising discrepancy. *Psychological Science*, 12, 141–147.
- Jenkins, H. M., & Ward, W. C. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs: General and Applied*, 79, 17.
- Johnson, S. P., Amso, D., & Slemmer, J. A. (2003). Development of object concepts in infancy: Evidence for early learning in an eye tracking paradigm. *Proceedings of the National Academy of Sciences*, 100, 10568–10573.
- Johnson, S. P., Bremner, J. G., Slater, A., Mason, U., Foster, & Cheshire, A. (2003). Infants' perception of object trajectories. *Child Development*, 74, 94–108.
- Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of sound patterns of words in fluent speech. *Cognitive Psychology*, 29, 1–23.
- Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. In B. A. Campbell, & R. M. Church (Eds.), *Punishment and aversive behavior*. New York: Appleton-Century-Crofts.\vedq11\
- Kirkham, N. Z., & Johnson, S. P. (2004). *Moving sounds: The role of inter-modal perception in solving the problem of occlusion*. Manuscript submitted for publication.\vedq12\
- Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy. *Cognition*, 83, B35–B42.
- Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2004). *Location, location, location: Development of spatiotemporal sequence learning in infancy*. Manuscript submitted for publication.\vedq13\
- Kruschke, J. K., & Blair, N. J. (2000). Blocking and backward blocking involve learned inattention. *Psychonomic Bulletin and Review*, 7, 636–645.
- Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science*, 218, 1138–1140.
- Lagnado, D. A., & Sloman, S. (2004). The advantage of timely intervention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 856–876.
- Leslie, A. M., & Keeble, S. (1987). Do 6-month-old infants perceive causality? *Cognition*, 25, 265–288.
- Lillard, A. S., & Witherington, D. C. (2004). Mothers' behavior modifications during pretense and their possible signal value for toddlers. *Developmental Psychology*, 40, 95–113.
- Madole, K. L., & Cohen, L. B. (1995). The role of object parts in infants' attention to form-function correlations. *Developmental Psychology*, 31, 637–648.
- Madole, K. L., & Oakes, L. M. (1999). Making sense of infant categorization: Stable processes and changing representations. *Developmental Review*, 19, 263–296.
- Meltzoff, A. N., Gopnik, A., & Repacholi, B. M. (1999). Toddlers' understanding of intentions, desires and emotions: Explorations of the dark ages. In P. D. Zelazo, J. W. Astington, & D. R. Olson (Eds.), *Developing theories of intentions* (pp. 17–46). Mahwah, NJ: Erlbaum.
- Michotte, A. (1963). *The perception of causality*. Oxford, England: Basic Books.
- Morronegiello, B. A., Fenwick, K. D., & Chance, G. (1998). Cross modal learning in newborn infants: Inferences about properties of auditory-visual events. *Infant Behavior and Development*, 21, 543–554.
- Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, 92, 289–316.
- Nazzi, T., & Gopnik, A. (2000). A shift in children's use of perceptual and causal cues to categorization. *Developmental Science*, 3, 389–396.
- Needham, A., & Baillargeon, R. (1993). Intuitions about support in 4.5-month-old infants. *Cognition*, 47, 121–148.
- Pearl, J. (2000). *Causality*. New York: Oxford University Press.
- Piaget, J. (1929). *The child's conception of the world*. London: Routledge & Kegan Paul.
- Rakison, D. H. (2004). Infants' sensitivity to correlations between static and dynamic features in a category context. *Journal of Experimental Child Psychology*, 89, 1–30.
- Reichenbach, H. (1956). *The direction of time*. Berkeley, CA: University of California Press.
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning*

II: Current theory and research (pp. 64–99). New York: Appleton-Century-Crofts.

- Richardson, D. C., & Kirkham, N. Z. (2004). Multimodal events and moving locations: Eye movements of adults and 6-month-olds reveal dynamic spatial indexing. *Journal of Experimental Psychology: General*, 133, 46–62.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926–1928.
- Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, 70, 27–52.
- Schult, C. A., & Wellman, H. M. (1997). Explaining human movements and actions: Children's understanding of the limits of psychological explanation. *Cognition*, 62, 291–324.
- Schulz, L. E., & Gopnik, A. (2004). Causal learning across domains. *Developmental Psychology*, 40, 162–176.
- Shanks, D. R. (1985). Forward and backward blocking in human contingency judgment. *Quarterly Journal of Experimental Psychology*, 37b, 1–21.
- Shanks, D. R. (1995). Is human learning rational? *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 48, 257–279.
- Shanks, D. R., & Dickinson, A. (1987). Associative accounts of causality judgment. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 21, pp. 229–261). San Diego, CA: Academic Press.
- Shultz, T. R. (1982). Rules of causal attribution. *Monographs of the Society for Research in Child Development*, 47, 1–51.
- Smith, L. B., & Heise, D. (1992). Perceptual similarity and conceptual structure. *Advances in Psychology*, 93, 1–21.
- Sobel, D. M. (2004). Exploring the coherence of young children's explanatory abilities: Evidence from generating counterfactuals. *British Journal of Developmental Psychology*, 22, 37–58.
- Sobel, D. M., & Kirkham, N. Z. (2005). *Blickets and babies: The development of causal reasoning in toddlers and infants*. Manuscript submitted for publication, Brown University.
- 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.
- Spelke, E. S. (1979). Perceiving bimodally specified events in infancy. *Developmental Psychology*, 15, 626–636.
- Spelke, E. S. (1981). The infant's acquisition of knowledge of bimodally specified events. *Journal of Experimental Child Psychology*, 31, 279–299.
- Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review*, 99, 605–632.
- Spelke, E. S., & Van de Walle, G. (1993). Perceiving and reasoning about objects: Insights from infants. In N. Eilan, R. A. McCarthy, & B. Brewer (Eds.), *Spatial representation: Problems in philosophy and psychology* (pp. 132–161). Oxford, England: Blackwell.
- Tenenbaum, J. B., & Griffiths, T. L. (2003). Theory-based causal inference. *Proceedings of the 14th Annual Conference on the Advances in Neural Information Processing Systems*.
- Tenenbaum, J. B., Sobel, D. M., & Gopnik, A. (2004). *Bayesian inference in causal learning from ambiguous data: Evidence from adults and children*. Manuscript submitted for publication, Massachusetts Institute of Technology.
- Uzgiris, I. C., & Hunt, J. M. V. (1975). *Assessment in infancy: Ordinal scales of psychological development*. Oxford, England: University of Illinois Press.
- Waldmann, M. R., & Hagmayer, Y. (2001). Estimating causal strength: The role of structural knowledge and processing effort. *Cognition*, 82, 27–58.
- Walker, A. S. (1982). Intermodal perception of expressive behaviors by human infants. *Journal of Experimental Child Psychology*, 33, 514–535.
- 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: Comparative and Physiological Psychology*, 51, 121–138.
- Wentworth, N., Haith, M. M., & Hood, R. (2002). Spatiotemporal regularity and interevent contingencies as information for infants' visual expectations. *Infancy*, 3, 303–321.
- Younger, B. A. (1990). Infants' detection of correlations among feature categories. *Child Development*, 61, 614–620.
- Younger, B. A., & Cohen, L. B. (1983). Infant perception of correlations among attributes. *Child Development*, 54, 858–869.
- Younger, B. A., & Cohen, L. B. (1986). Developmental change in infants' perception of correlations among attributes. *Child Development*, 57, 803–815.
- Younger, B. A., & Gottlieb, S. (1988). Development of categorization skills: Changes in the nature or structure of infant form categories? *Developmental Psychology*, 24, 611–619.

\edq16\

\edq17\

\edq14\

\edq15\

- \edq1\Cross reference to chapter 16.
- \edq2\Cross reference to chapter 5. Is this the correct chapter?
- \edq3\The term backwards changed to backward throughout per *MW Collegiate*. Okay?
- \edq4\Cross reference to this volume. Clarify throughout whether you mean Tenenbaum, Griffiths, & Niyogi, chapter 19, or Griffiths & Tenenbaum, chapter 20.
- \edq5\Cross reference; provide chapter and correct authors.
- \edq6\Provide Tenenbaum et al., 2005, reference. If not, change to 2004 throughout.
- \edq7\Cross reference; provide chapter and correct authors.
- \edq8\Provide Griffiths and Sobel, in preparation.
- \edq9\Provide editor's name and inclusive pages for Christiansen & Dale, 2001.
- \edq10\Update Griffiths & Tenenbaum, in press.
- \edq11\Provide inclusive pages for Kamin, 1969.
- \edq12\Update Kirkham & Johnson, 2004.
- \edq13\Update Kirkham et al., 2004.
- \edq14\Provide inclusive pages for Smith & Heise, 1992.
- \edq15\Update Sobel & Kirkham, 2005.
- \edq16\Follow APA guidelines for proceedings for Tenenbaum & Griffiths, 2003.
- \edq17\Update Tenenbaum et al., 2004.