In P. Zelazo (ed). The Oxford Handbook of Developmental Psychology. New York: Oxford University Press. (In press).

Causality

Alison Gopnik

Dept. of Psychology

University of California at Berkeley

This research and the preparation of this chapter were supported by the James McDonnell Foundation Causal Learning Collaborative Grant and by NSF grant BCS 1023875. I thank Adrienne Wente for assistance.

Authors address:

Alison Gopnik

Dept. of Psychology

University of California at Berkeley

Berkeley, CA 94720

gopnik@berkeley.edu

Keywords: Causal learning, conceptual development, intuitive theories, probabilistic models, causal Bayes nets, Bayesian learning

Abstract

Causal knowledge is central to children's understanding of the world but there have been many different conceptions of how causal knowledge is represented and how it is learned. I outline four different approaches to causality that have influenced developmental research, stressing agency, mechanism, association and probabilistic models, and give examples of theory and research in each area. I focus in more detail on the most recent probabilistic model accounts which stress the role of causation in inferences about possibilities. These include potential interventions to change the world and counterfactuals about alternate ways the world might be. These accounts also employ computational ideas about causal Bayes nets and Bayesian learning, and these ideas are outlined.

Causality

"Causality is the cement of the universe", said David Hume (1740). Causal knowledge is arguably the most significant kind of knowledge we can have, uniting our apparently disparate experiences, allowing us to make new predictions about the events we have yet to experience, and most significantly, allowing us to imagine new possibilities and intervene in the world to reshape it in significant ways. We could argue,

in fact, that causal knowledge underlies our distinctively human ability to control and construct our environment.

And yet there is something elusive about this most central kind of human knowledge. Philosophers and psychologists have given a wide array of different characterizations of causality, emphasizing strikingly different kinds of phenomena. These include psychological phenomena of agency, action and outcome, physical and spatial phenomena of movement, contact and mechanism, patterns of association and co-variation, and more abstract relationships of probability, intervention and counterfactual dependence. Indeed, this proliferation of competing conceptions of causation led the philosopher Bertrand Russell (1912) to declare that causality should be ruled out of philosophical discussion altogether "The law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm".

This range of philosophical conceptions of causation is mirrored by the very wide range of developmental investigations that might be described as studies of causal development. These include the early Piagetian conception of causation in terns of agency, a number of investigations in the eighties that focused on the role of spatial configuration and physical mechanisms in causal understanding, and the associationist conceptions that were the underpinning of classical learning theory and later connectionism. Most recently, there has been a renewed interest in probabilistic,

interventionist and counterfactual accounts of causation, especially in the context of computational work on probabilistic models.

Moreover, in addition to studies framed in terms of the development of an understanding of causation itself, there have been many studies of the development of different domains of causal knowledge, including the details of everyday physics, psychology and biology. On this conception however, the study of causal development would overlap with a great deal of cognitive and conceptual development, including a large number of the other chapters in this handbook! In particular, almost all of the studies of children's intuitive theories of the world, including the enormous body of work on "theory of mind", are examples of the development of causal knowledge in this sense.

Consider the concepts and laws of intuitive theories. These might include the principle that beliefs and desires lead to action or that biological essences are responsible for the correlations among features of animals or that objects convey momentum to other objects. All of these principles might be characterized as causal generalizations and the central concepts that are involved in intuitive theories might be seen as particularly causal in nature. Indeed, causal structure has often been seen as one of the distinctive features of intuitive theories (Gopnik & Melzoff, 1997). From this perspective only a few aspects of conceptual development, the development of spatial and numerical knowledge, for example, would NOT be causal.

Within the literature on children's intuitive theories, however, most of the emphasis has been on characterizing the details of children's causal knowledge of particular domains and the changes in that knowledge as children grow older (Flavell, 1999; Gelman 2003; Gelman & Ramen 2002; Wellman & Inagaki, 1997). Infants, for example, seem to understand how occlusion works before they understand containment (Baillargeon, 2008) and preschoolers understand inheritance before they understand life and death (Gelman, 2003), and desire and perception before they understand belief (Wellman, 1992; Wellman & Liu, 2004). There has been relatively little attention in this literature to the very nature of causation itself or what it means to say that intuitive theories involve causal claims. And there has been relatively little work investigating the reasons that these theory changes take place and the techniques that children might use to infer patterns of causal structure from evidence.

Of course, it is possible that the causal schemas and principles underlying these different domains are indeed profoundly different, and that there is no unifying set of representations or learning procedures that apply to them all. But it is also possible that by understanding causal knowledge and causal inference more generally we might be able to go beyond charting the details of children's causal knowledge and say something more general about the nature of that knowledge, or the processes by which it is learned.

In this chapter I will discuss several different conceptions of causality in philosophy and psychology, outline some of the historical work on the development of causal understanding that reflects each of these conceptions, and then focus on contemporary

work in each area, much of it still in progress. In particular I will give a fairly extensive account of the most recent probabilistic models approach, partly because that approach has proved particularly fruitful, but also because the computational character of much of this work may make it initially seem rather opaque to most developmentalists (including the present author!). Although mathematical details always seem formidable, the conceptual framework of probabilistic models is actually straightforward, or at least I will try to make it that way. At the end I will return to the question of children's intuitive theories and the relation between theory development and causal development more generally. And, finally, I will describe some unanswered questions that present exciting possibilities for future empirical work.

Conceptions of Causality

There are (at least) four very different, and even competing approaches to causation that come out of the philosophical and psychological literature, which I'll call agency, mechanism, association, and probabilistic model accounts, reflecting the variety of our everyday intuitions about what causality is all about.

Agency

As always in cognitive development, Piaget was the first scientist to think seriously about the development of causal knowledge. And, as always, Piaget's ideas on causation are complex, subtle and extremely insightful in guiding us towards interesting developmental

phenomena -- and have turned out to be empirically wrong. Piaget's central idea was that our understanding of causation had its roots in children's experience of their own agency and control over the world (Piaget, 1929). On this view the canonical example of a causal connection was the link between our intention to act to bring about a goal and the fulfillment of that intention. Because children's notion of causation had its origins in this egocentric notion of agency, there were important limitations on the kinds of causal inferences that children could make.

This, in turn, led Piaget to argue that preschooler's conception of the world was "precausal", at least compared to the conceptions of older children and adults. In particular, when Piaget asked children to explain such events as the succession of night and day, using his "clinical interview" technique, children often produced "animistic" responses. That is, they answered in term of their own desires or goals or intentions rather than in terms of the objective causal relations in the world. Children might, for example, say that there is night so that we can sleep. Although this was a mistake, it was an illuminating mistake in so far as it suggested where our ideas about causation came from. For Piaget causal development involved a gradual progression away from the early agent centered view of causation to a more objective "decentered" view (Piaget, 1929).

Interestingly, an opposing though historically contemporaneous theory of learning, the view of operant conditioning, actually suggested a similar view of causality, at least implicitly. In operant conditioning, after all, the basic principle is that an animal's intentional actions are shaped by the immediate outcomes of those actions. One might at

least interpret this as a claim about the animal's understanding of causality – an understanding that would privilege the relations between the animal's own actions and their outcomes. Like Piaget's infants, the rats in a Skinner box might be construed as acting as if they thought that their actions caused relevant outcomes, and therefore repeated actions that led to favorable consequences, but did not extend their causal inferences beyond this narrow scenario (Skinner, 1938).

In concert with both the Piagetian and operant conditioning ideas, there was a large body of work starting in the eighties, showing that even very young infants will shape their future actions in the light of the outcomes of their earlier actions. In particular, Rovee-Collier and her colleagues, among others, showed that infants would very effectively learn the contingent consequences of their own actions (Mast et al, 1980; Rovee-Collier, & Sullivan 1980; Rovee-Collier, & Sullivan et al 1980; Watson & Ramey, 1987). Even three-month-olds, for example, would quickly learn that kicking a foot with a ribbon tied to it would lead a mobile attached to that ribbon to move. Unlike classical operant conditioning, however, the infants in Rovee-Colliers studies did not depend on reward contingencies, but simply learned those relations and used them to shape future actions. For these infants learning causal relations seemed to be its own reward, and we might think of this more as a kind of trial-and-error causal learning than as classical operant conditioning.

Some philosophers and psychologists have tried to argue for the agent-based view more generally, as an account of general causal cognition and not just cognition in infancy

(White, 1995; Menzies & Price, 1993). However, Piaget is right in saying that though this view may provide the origins of an understanding of causation it cannot account for the wide range of causal inferences we are capable of as adults, including many that are not the result of human agency at all. Still, as we will see, Piaget's genuine insights about the link between agency and causation inform many of the ideas that have been recently developed in "interventionist" accounts of causation within a probabilistic modeling framework.

2) Mechanism

Mechanism accounts stress the spatio-temporal aspects of causal events. The quintessential example of causation in these accounts is what is sometimes called "billiard ball" causation. If you see one ball knock into another and the second ball move off, it is very difficult not to conclude that the first ball caused the second ball's movement. Back in the seventies Michotte (1962) famously showed the strength of this impression by giving people films of various collision events and showing that they judged some but not others to be causal, based on their spatio-temporal properties. More recently Scholl has taken up Michotte's idea and his psychophysical methodology and has shown that adult perceivers are extremely sensitive to fine details of the course of the object's movement and the nature of the collision in making these judgments (Scholl & Nakayama, 2002).

This sort of spatiotemporal and mechanical transmission was the focus of what we might think of as the second wave of studies of causality, studies that were the first empirical corrective to the Piagetian tradition and its claim of causal deficits in early childhood. In two parallel sets of studies Rochel Gelman and colleagues and Thomas Shultz showed that, in fact, and contra Piaget, even young preschoolers could understand complex cases of physical causation and could make appropriate inferences and draw correct conclusions. Three and four-year-old children in these experiments were exposed to rather complex novel physical machines working on various kinds of movement, transmission and collision principles. Children were excellent at understanding, and, in particular, predicting the behavior of these machines, even when the mechanisms were quite independent of their own actions and interventions. For example, Bullock et al. (1982) showed preschoolers an apparatus that involved a long rod that pushed an initial domino like object, leading to a causal sequence the ended with a rabbit tumbling over (see Fig. 1) When children saw only the first and last part of the sequence they could nevertheless infer the intermediary movements that were necessary to cause the final event involving the rabbit. Similarly, in a series of experiments, Shultz (1982) demonstrated that in their causal judgments, preschoolers privilege evidence for spatially continuous processes compatible with the transmission of energy, over evidence for covariation. Preschoolers inferred, for instance, that a tuning fork whose vibrations were not obstructed was more likely to produce a sound than a tuning fork whose vibrations were blocked, even when the effect immediately followed an intervention on the latter and followed the former only after a delay.

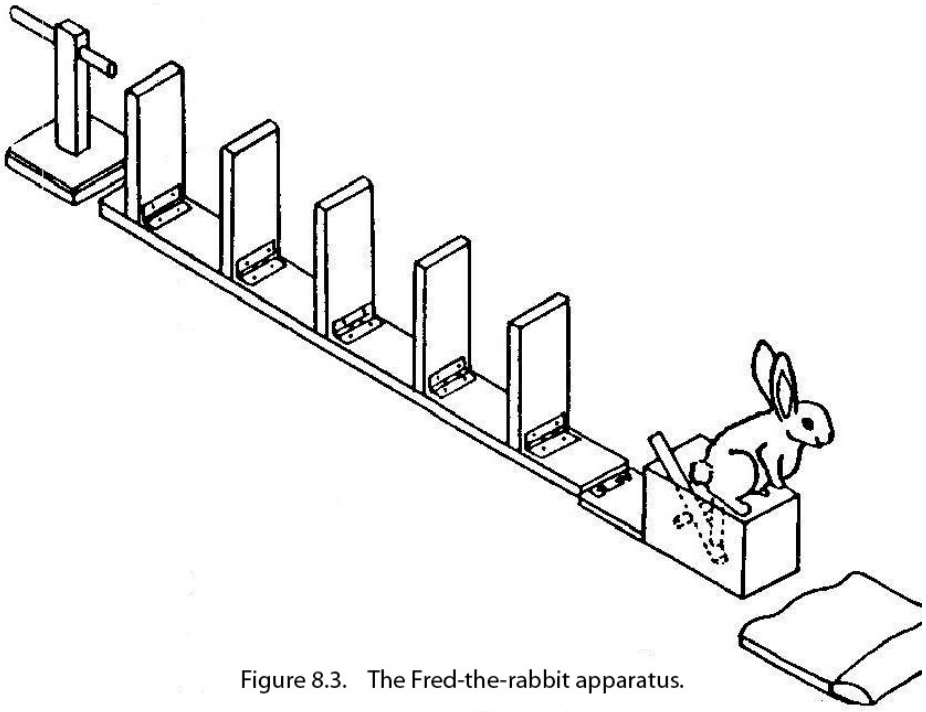


Figure 8.3. The Fred-the-rabbit apparatus.

These results led to the suspicion that Piaget's original results stemmed from the abstract and unfamiliar nature of the causal questions he asked. When children were asked causal questions about familiar kinds of physical causal relations they appeared to be much more competent.

Going beyond Michotte, the children in these studies also seemed to infer physical causal mechanisms even when there was no direct perceptual evidence for such transmission, when the causal sequence was hidden from view, for example, or was even invisible, as in the tuning fork case. More recently Schlottman (1999) has obtained similar results and

has shown that, in preschoolers, knowledge about physical mechanism can trump perceptual impressions.

Alan Leslie extended the Michottean idea to an even younger age range. In studies of six-month-olds he used a looking-time procedure to explore the babies understanding of physical causal events. He habituated babies to either a contact or non-contact event (A bumps B and B launches or A moves towards B without contact and B launches). Then he reversed the time order of that event (So that babies saw B launch first followed by A's movement). These babies looked longer at the reversal in the contact case but not the non-contact case. This suggests that the infants at least treated causal and non-causal sequences differently (Leslie & Keeble 1987). Infants similarly made differentiations among causal events involving human hands. When babies saw a hand make contact with an object and then move off with it they looked less long than when the hand failed to make contact with the object and hand and object moved off together. They did not discriminate in this way between two sequences that involved objects, rather than hands, making contact.

Oakes and Cohen (1990) replicated Leslie's results with his materials for seven-month-olds but also found that the infants initially behaved in this way for a relatively narrow set of perceptually defined events, involving quite particular movement trajectories and objects, echoing Scholl's psychophysical results with adults. For example, when six-month-old infants saw toy vehicles making contact, rather than the simple blocks in

Leslie's design, they did not discriminate causal from non-causal events, though 10-month-old infants did. Over the course of the first eighteen months, however, infants gradually became more and more adept at making such inferences across a wider range of movement events and objects.

In the philosophical literature writers like Dowe (2007) and Salmon (1998) have argued that causality requires a "continuous process of spatio-temporal transmission", some constant process by which the effect can be traced to contact with the cause. Such an account does seem to capture some of our basic intuitions about causality. It may explain our intuitive dissatisfaction with scientific accounts that seem to allow for action at a distance, such as quantum theory, or, for that matter, Newtonian gravitation, and our tendency to infer invisible forces when we face a causal connection that does seem to involve action at a distance.

On the other hand, most of the causal relations we learn about in everyday life do not depend on this sort of perceptual and spatio-temporal information. From inferring that your desire for milk makes you open the fridge, to inferring that an essence in a seed makes it grow into a flower, to inferring that cigarette smoking causes cancer, we are able to make a causal inference long before we have a detailed story about spatio-temporal transmission, if we ever do.

In fact, Rozenblit and Keil (2002) have shown empirically that adults have what he calls an "illusion of explanatory depth". Often adults seem to think that they have mastered a

causal understanding of some phenomenon, from helicopters to zippers, although their understanding of the physical mechanisms that underlie that phenomenon is, at best, partial and shallow.

For some cases, certain cases of biological or psychological causation like the seeds and desires, for instance, it might be that there are specific modules or innate schemas, analogous to a Michottean schema for physical causation. Something like this has been suggested by Leslie (1994). However, we also clearly learn a wide range of new causal relations that violate the rules of any plausible innate modules, learning, for example that a remote control can make a TV turn on or that anxiety can cause stomach aches, or for that matter, that smoking causes cancer.

Philosophers have also pointed out that some types of causal influence simply have no associated pattern of spatio-temporal transmission at all (Hitchcock, 2007). For example, consider cases of causation by omission. Suppose a watchman falls asleep, allowing a burglar to rob the warehouse, or an oxygen cutoff valve fails to operate, allowing a fire to ignite. In both cases it seems natural to say that what the watchman or valve did NOT do caused the disaster, in spite of the fact that there is no spatiotemporal process connecting the two events.

The philosophical view, and indeed our ordinary intuitions, suggest that cases of spatio-temporal contact and physical transmission, like cases of goal-directed action, may indeed be particularly exemplary cases of causation, but they do not exhaust the kinds of

causal relations we understand as adults and perhaps as children. It remains to be seen of course, whether these might be the only initial types of causal understanding and other causal abilities gradually develop from them.

Association

Another set of approaches to causation, dating back to David Hume (1740), suggest that causation is nothing but the association of two events. Hume famously argued that when two events always occur together we come to predict that one thing will happen when the other does. This impression of association, and the predictions and inferences it leads to, constitute our experience of causality. Although Hume himself is somewhat ambiguous on this point, this view has often been taken to imply that causality is an illusion -- "nothing but" association and covariation.

Just as we might think that there is an implicit agency theory of causation in operant conditioning we might also think that there is an implicit associationist theory of causation in classical or Pavlovian conditioning. In fact, one of the theoretical crown jewels of the study of conditioning, the Rescorla-Wagner equation, can be interpreted as an algorithm for drawing causal conclusions from covariational data (Danks, 2003). Again, though, the arguments are often phrased the other way around. Researchers in the animal learning tradition, like Dickinson or Shank, for example, have argued that causal inference in humans is really nothing but associationist or Pavolvian learning (Shanks, 1995; Shanks & Dickinson, 1987).

In development, we know that infants are capable of Pavlovian or classical conditioning as well as operant conditioning (Blass et al, 1984). Moreover, connectionist computational accounts of development have taken the associationist picture much further than the initial accounts of those like Rescorla and Wagner. These accounts allow for complex patterns of association building up over time, and altering the strengths of the predictive relationship between pieces of evidence. Recently, Rogers and McLelland (2004) have used such a model to account for early causal understanding in three to five year olds, and have shown that gradual associations over many trials can lead to the same pattern of predictions about future events that we see in children's causal inference. Oakes and Cohen (1990) make a similar argument about infants' gradually broadening inferences in the Leslie paradigm. Indeed one could argue that even the classical Michottean phenomena can be explained in terms of covariation. Since certain types of physical events are followed so often by particular types of physical outcomes, we come to think of those events as causally related.

In philosophy, the equivalent of the modern associationist account has been accounts defining causality in terms of probability. Reichenbach (1956) for example, suggested that we could define a causal relation as one in which the probability of the cause and the probability of the effect were related to one another in systematic ways. In particular, Reichenback argued that events could be construed as causal when they were correlated in ways that allowed for "screening off", an idea I will elaborate in more detail below.

Reichenbach's notion of "screening-off is actually rather similar to the relation between probability and causality captured by the Rescorla-Wagner equation.

If agency and mechanism accounts of causation appear to be too narrow to capture all the events we call causal, however, associationist accounts appear to be too wide. After all, every first year statistics student learns that correlation does not imply causation. Consider, for example, the contrast between our knowledge of the causal relation between smoking and getting lung cancer and our knowledge of the correlational relation between having yellow fingers and getting lung cancer. Since smoking leads to yellow fingers, yellow fingers will indeed be correlated with cancer, and may even be an excellent predictor of lung cancer. But we feel that it is smoking that actually causes cancer while yellow fingers do not. So many correlational relations do not seem to imply causation. Mere association is inadequate to account for our causal intuitions. What is the extra element that turns covariation into causation?

Probabilistic models, Bayes nets and the interventionist account

The most recent wave of interest in causal development comes along with a revival of interest in causality in philosophy and computer science. This new approach to causality has come to dominate discussions of causal inference in philosophy, computer science and statistics and has become increasingly influential as an account of causal knowledge and learning in both adults and children. The approach has both a particular philosophical and conceptual foundation – the interventionist account of causation -- and a particular

formal and computational instantiation – probabilistic graphical causal models, aka causal Bayes nets. It combines ideas about agency, probability, and association into a single unified and formally coherent theory. Notably missing from this list, of course, are ideas involving spatio-temporal features and mechanism and I will return to this point later on.

Counterfactuals and interventions

Let's go back to the question of how correlation differs from causation. The philosopher David Lewis (1973) first suggested that the relevant difference between correlation and causation involved counterfactual inferences. Causal claims, as opposed to mere correlations, will also support counterfactual inferences. If smoking does cause cancer then if we had banned cigarettes in the past there would be less lung cancer now – if John had given up smoking at 20 he would have been less likely to get cancer at 70, if a group of lab mice were to inhale smoke the probability that they would get cancer would go up and so on. None of these inferences will hold for the mere correlation between yellow fingers and cancer. Instituting a hand-washing campaign would have had no effect on cancer, nor would painting the mice's paws with nicotine extract. This account also seems to capture some of our intuitions about the omission cases that are troubling for the mechanistic accounts – we feel that the watchman's nap caused the robbery because had it been otherwise the robbery would not have taken place.

The difficulty with the counterfactual account from a psychological, developmental and even evolutionary point of view is that we can't have any direct evidence for counterfactual claims, which are, after all, precisely things that haven't happened. Equally, it is difficult to see how accurate inferences about non-existent events could lead to better reproductive success! So though counterfactual inferences may be a consequence of causal knowledge they leave open the question of how that causal knowledge could be learned or understood psychologically. Moreover, developmentally, counterfactual understanding seems to be a later acquisition than at least some types of causal understanding. There is strong recent evidence that children understand counterfactuals earlier than we once thought (Harris, German, & Mills,1996) and that counterfactual understanding and causal understanding are linked (Sobel, 2004). However, it is difficult to tell how we could even test for counterfactual understanding without the use of language.

The Interventionist Theory

The interventionist theory of causation also proposes that a type of conditional inference is central to causal claims, one that goes beyond simple prediction. The interventionist account, however, stresses future hypothetical inferences rather than the typical "had it been otherwise" past counterfactuals of Lewis' account. These are inferences about what could happen in the future rather than what could have happened in the past. The interventionist theory has been most clearly formulated by James Woodward (2003) in

his book Making Things Happen, though it is also implicit in formal ideas about manipulation in Pearl (2000) and Spirtes et al (2000).

Let's go back to the example of the relation between smoking and cancer, and yellow fingers and cancer. We could articulate the difference in terms of counterfactuals, what could have happened in the past, for example, but we can also articulate it in terms of policies for the future. If we believe that smoking causes cancer we can intervene, that is, actively set about influencing the smoking rate, in order to change the rate of cancer. We don't think that a similar intervention on yellow fingers is warranted despite the correlation. So, roughly speaking, we can define what we mean by saying that A is directly causally related to B, by saying that, other things equal, if we intervened to make a change in the probability of A ("wiggling A" is often the technical philosophical term) then there would be a resulting change in the probability of B.

This definition seems to capture the same set of intuitions that the counterfactual definition captures. But it leads to a more evolutionarily and developmentally plausible story. An ability to plan, and to act in order to bring about particular outcomes, is, after all, just the sort of thing that would lead to evolutionary advantages, and intentional action is apparent even in infancy and certainly well before sophisticated counterfactual language is possible.

Like the Piagetian agency theory this account connects to ideas about human action. The account differs from the agency view in two important ways though. First, the agency

view stresses the causal relation between the intention to act and the immediate outcome of that intention. The interventionist view, in contrast, stresses the relation between the "down-stream" events following that action. The picture is that there is some objective relation between the cause and effect out there in the external world. The nature of that relationship is such that a human intervention on the cause could lead to a change in the effect, at least, in principle. This is why experiments are the best and most direct way to discover causal structure – in an experiment we perform an intervention on A and then observe whether there is a corresponding change in B. If there is, we conclude that A and B are causally related.

Second, at least in the philosophical and formal version, an intervention is not just any goal-directed human action. An intervention is an action with very special features. Intuitively, you can understand these special features by thinking about the interventions you would perform in an ideal experiment. Consider an experimental drug trial, for example, where I am trying to test whether administering a new drug will cause a patient to improve. On the interventionist view, I can conclude that the drug does cause the effect if, controlling all the other factors, I act to change the drug level in the patient and there is a change in the patient's symptoms.

However, this will only be true if my action has particular features. My decision to give one group of people a treatment rather than another should not be influenced by the outcomes, it should, in fact, be randomly determined. It should ensure that the treatment group does actually take the drug. And the treatment should only affect the patient

through the influence of the drug itself and not through some other route. To count as an intervention my action must be what statisticians call "exogenous" – that is it must not be influenced by the causal system that is being intervened on. It must affect the variable it intervenes on, and it must only affect the "downstream" variables through its influence on that variable.

Human actions often have these features On the other hand, they don't always have these features, as scientists know to their peril. In cases of experimenter bias, the outcome of one experimental trial of the drug, for example, might subtly influence the way I intervene on the next trial. In cases of non-compliance, my intervention might not affect the cause at all – I might tell the patients to take the drug but they might fail to do so. In placebo effects, my intervention might cause a change in the symptoms, but it would be because any treatment improves the symptoms directly rather than because this particular treatment does. In all these cases, the action would not be an intervention on the interventionist view, and so inferences about causation would not be valid.

On the other hand, on the philosophical view, an intervention need not actually be an intentional human action designed to bring about the effect. We might have what statisticians call "exogenous variables" or "natural experiments". Natural experiments can be treated as interventions even if they aren't intentional actions designed to bring about the relevant effect. For example, consider the question of whether attending kindergarten causes later cognitive achievement. We might discover that kindergarten attendance is correlated with achievement but that by itself would fail to tell us whether

there was a causal relationship between the two – perhaps brighter children are simply more likely to be enrolled in kindergarten. We might not, for ethical or practical reasons, be able to intervene to randomly put some children in kindergarten and not others. However, we can exploit the fact that many school systems have arbitrary age cut-offs for kindergarten attendance, so that children who are born on Dec 31 will be treated differently than those born on Jan 1. Although this was not intended to influence cognitive achievement we can treat it as a kind of intervention – it is not influenced by achievement, it determines whether or not children are in kindergarten, and it doesn't itself influence achievement by other means. If children who were born on the 31st do better than those born on the first we can conclude that assignment to kindergarten did causally influence achievement.

The interventionist account is intended to be what philosophers call "normative" – it's supposed to tell us the right way of making true causal inferences. To serve this function the account involves many technical and philosophical complexities and subtleties, like the subtleties involved in exactly defining an intervention. Of course, an empirical question of some interest is whether children or even adults recognize these subtleties of causal inference or whether they simply treat all human goal directed actions as interventions, and draw causal conclusions on that basis, sometimes incorrectly.

Independent of those subtleties, however, the interventionist idea does seem very plausible psychologically. It's hard to think of examples of a relationship that most

humans would think of as causal and that did not respond to interventions in this way, at least, in principle. It is true, almost by definition, of the relation between our intentions to act and the production of those actions. It is true of all the Michottean physical cases. We certainly think that if we made one ball hit a second ball, the second ball would roll, while that would not be true if we simply pushed one ball near the other. But it is also true of the myriad "explanatorily shallow" cases in which we have no understanding of the spatio-temporal mechanism, but we nevertheless feel that there is a causal link. One need not understand the mechanism of zippers, for example, to know that the intervention of pulling on the zipper tag will cause the dress to open. And knowledge of such interventions is precisely what contributes to evolutionary and reproductive success (as the zipper case illustrates, of course).

The interventionist view also gives us a way of psychologically discriminating mere cases of correlation from cases of causality. Our intuition that yellow fingers do not cause cancer, though they may predict it, can be cashed out in terms of the interventions we could perform to make certain outcomes more likely – finger-washing interventions would have no effect. Associations that do not support interventions are not causal. So the interventionist account gets at the Goldilocksian middle ground between the too small agency and mechanism accounts and the too big associationist ones.

Probabilistic and models and Bayesian inference

The interventionist account has been closely connected to a set of formal and computational ideas that have emerged in the last 15 years or so in computer science and philosophy. These are ideas about "probabilistic models". Causal Bayes nets or causal graphical models apply probabilistic modelling to causal knowledge and learning, in particular. But probabilistic models have also been applied to many other problems in machine learning and there has been a movement applying probabilistic models to many other problems in cognitive science and cognitive development. Probabilistic models are starting to be used to account not only for causal cognition and development but for many other types of cognition and development – ranging from vision (e.g. Han & Zhu 2005), to language (e,g. Xu & Tenenbaum), to hierarchical categorization (e.g. Kemp & Tenenbaum 2009) (see Griffiths, et al, 2010 for a summary and review). I will outline some of the general ideas behind proabilistic models first here and then provide more details about causal Bayes nets, in particular.

The debates in the cognitive science of causality typify a broader historical debate in cognitive science in general. Throughout the history of cognitive science there has been a tension between two fundamental approaches to knowledge. On the one hand it seems that children and adults have abstract, hierarchical, highly structured and accurate representations of the world – representations that support a very wide range of predictions and inferences. On the other hand, it seems that we have to somehow learn that abstract structure from our very concrete, immediate, particular experiences. It has traditionally seemed very difficult to bridge that gap.

The opposing traditions in cognitive science reflect commitments to one horn or the other of this dilemma. The nativist and rationalist tradition accepts the idea that knowledge is abstract, hierarchical and structured but denies that that structure is learned from contingent, concrete and probabilistic evidence. The empiricist and associationist tradition claims that knowledge involves the gradual accumulation of concrete patterns of evidence but denies that there is abstract hierarchical structure. In the causal case, in particular, the Michottean tradition advances the rationalist argument – the detailed structure of mechanical causation is innately given - while the associationist tradition argues the empiricist claim – causation just is correlation.

Developmentalists, who empirically seem to see both structure and learning, have long searched for a way out of this dilemma. Piagetian constructivism proposed that the gap should be bridged but failed to provide a concrete way of doing so. It is a cliché in developmental psychology that the truth must lie in the interaction between nature and nurture, but it has been hard to say more than that. Probabilistic models allow us to articulate not just a wishy-washy middle ground, but a genuinely new way of solving the problem.

The breakthrough came from describing kinds of abstract structure that are related to probabilistic evidence in very predictable and coherent ways. Imagine that there is some real structure in the world – a three dimensional object, a grammar, a hierarchy or a causal relationship. That structure gives rise to some patterns of observable evidence rather than others – a particular set of retinal images, or spoken sentences or inclusion

relationships or outcomes of interventions. That grammar or hierarchy or spatial or causal structure can be mathematically represented, by a tree diagram or a three-d map or a causal graph, in a way that will allow you to generate the evidence. You could think of such a representation as a hypothesis about what the actual structure is like. You can predict what evidence should follow given that hypothesis, and so make many new inferences. If the hypothesis is correct, then these inferences will turn out to be right.

This systematic link between structure and evidence also allows you to reverse the process and to infer the nature of the structure from the evidence it generates. Vision scientists talk about this as "solving the inverse problem" (Palmer, 1999). Solving the inverse problem lets you learn about the world from evidence. It lets you decide which tree or map or graph is the best representation of the world outside.

The idea that mental models of the structure of the world generate predictions, and that we can invert that process to learn the structure from evidence, is not itself new. It is the basic model underlying both the cognitive science of vision and of language. In vision science we assume that the world has a three-dimensional structure that generates the retinal images we see. In linguistics we assume that structured grammars generate the sentences we hear. In both cases the child must somehow work backward from images and sentences to discover objects and grammar, and to construct accurate maps and trees.

The big recent advance has been integrating ideas about probability into that basic model. If you think of these mental models as logical or deterministic systems the inverse

problem becomes extremely difficult, if not impossible, to solve, and that has led to nativist conclusions (eg Chomsky, 2006, Pinker, 1996). Integrating probability theory makes the learning problem more tractable.

Many probabilistic models use Bayesian inference and learning techniques as a very general way to approach the inverse problem (though there are also non –Bayesian methods). Bayes' rule (in one version below) is a simple formula that relates the probability that a hypothesized structure generated the pattern of the evidence that you see, P (H/E), to the probability of that evidence given the hypothesis P (E/H), combined with the probability of the evidence itself P (E) and your initial estimate of the probability of the structure P (H). Each part of this formula has a conventional name P (B) is the "prior", your initial belief in the hypothesis, P (E/H) is the "likelihood", and P (H/E) – is the "posterior" -- your belief in the hypothesis after you've considered the evidence.

Bayes rule

P (H/E) = (P(E/H) (P(H)) % P(E)

Bayesian reasoning is probabilistic in several ways. The relations that are represented in the hypotheses may themselves be indeterminate. For example, a grammar might propose that an article will usually but not always be followed by a noun, a visual algorithm might propose that an object edge will usually but not always be continuous and a causal generalization may say that the presence of a desire will make a goal that fulfills that desire more likely but not certain.

Second, the relation between the hypothesis and the evidence may be probabilistic. A subject in a grammar may often but not always lead to the production of a noun, an edge may often but not always lead to a contour, and a desire may often but not always lead to an action.

Finally, rather than simply generating a yes or no deterministic decision about whether a particular hypothesis is true, the Bayesian learning algorithms simply say that one underlying structure is more probable than another. Often, in fact, usually, there are many spatial or causal structures, grammars, or hierarchies that could, in principle, produce a particular pattern of evidence. The structure is "underdetermined" by the evidence. This is the "poverty of the stimulus argument" that led Chomsky and others in the nativist tradition to argue that the structure must be innately given. But while many structures may be possible, some of those structures are going to be more likely than others. Bayesian methods give you a way of determining the probability of possibilities. They tell you whether one hypothesis is more likely than another.

Moreover, Bayesian reasoning provides a way to balance your prior knowledge – the assumptions about the world that you start with – with the force of new evidence. Bayes' rule integrates the prior probability of a hypothesis, the probability that it was true before you saw the evidence, with the probability of the evidence given the hypothesis. To take a classic example, consider the relation between a positive mammogram, E, and having cancer, H. The likelihood (P E/H) might be quite high, if you have cancer you are very likely to have a positive test. But the prior (P H) might be quite low, you might be very unlikely to actually have cancer and other factors might also lead to a positive mammogram. So the posterior, (P H/E) the probability that if you have a positive test you have cancer, might also be relatively low.

If a hypothesis is highly probable to begin with, it will take more evidence to overturn it. Eventually though enough evidence could lead you to accept even an initially very unlikely idea, many positive mammograms and other evidence might indeed lead to the conclusion that you had cancer. This gives Bayesian reasoning a characteristic combination of stability and flexibility.

Probabilistic models were originally articulated as ideal rational models for learning. Like signal detection theory or ideal observer theory in vision they tell us how a system could learn best, in principle. In fact, Bayesian ideas first emerged in the philosophy of science as a way of saying how scientists ought to react to new evidence, not necessarily predicting how they did react. However, signal detection theory and ideal observer theory can help us make hypotheses about how evolution actually did shape the visual system.

In the same way probabilistic models can help us formulate hypotheses about how evolution shaped human learning capacities. By using a kind of “reverse engineering” we can compare an ideal learning machine to the learning machine in our skulls.

These rational probabilistic models have both attractions and limitations as theories of the actual representations and learning mechanisms of cognitive development. One attraction is that they propose a way that, at least in principle, a child could genuinely move from one structured hypothesis to another very different hypothesis based on the patterns of evidence they saw. Children need not merely fiddle with the details of an innately determined structure or simply accumulate more and more evidence. They could genuinely learn something new.

The probabilistic character of the inferences also seems to map nicely on to the often gradual and piece-meal way that development proceeds – a common emphasis of empiricists. At the same time, the generative power of structured hypotheses can help explain the power and generality of children’s inferences – a common emphasis of nativists. And the integration of prior knowledge and new evidence is just what Piaget had in mind when he talked about assimilation and accomodation.

The drawback of probabilistic models is what is called the search problem. Bayesian reasoning gives you a way to evaluate a particular hypothesis, given a particular pattern of evidence. However, you still have to decide which hypothesis to evaluate. A very large number of hypotheses might be compatible with some particular pattern of evidence, and

the child obviously won't be able to go through and enumerate the probability of each one. Searching through all the possible hypotheses rapidly becomes intractable and computer scientists have had to devise ways to rationally sample just a few hypotheses at a time. However, within machine learning there are ways of at least approaching these problems, approximating an ideal search, and coming to progressively more accurate conclusions. We are beginning to explore whether children's minds may be doing something similar.

Causal Bayes Nets. Bayes' rule is only useful if you can say quite precisely what your hypothesis is, and say quite systematically how it generates evidence. Causal directed graphical models, or causal Bayes nets, do this very effectively for representations of causal structure. Causal Bayes nets have been developed in the philosophy of science and statistical literature over the last fifteen years (Glymour 2001; Pearl, 2000; Spirtes et al. 2000.) Scientists seem to infer theories about the causal structure of the world from patterns of evidence. But philosophers of science found it very difficult to explain how these inferences are possible. Although classical logic could provide a formal account of deductive inferences, it was much more difficult to provide an inductive logic – an account of how evidence could confirm theories. One reason is that deductive logic deals in certainties but inductive inference is always a matter of probabilities – acquiring more evidence for a hypothesis makes the hypothesis more likely, but there is always the possibility that it will be overturned. An even more difficult question was what philosophers of science called "the logic of discovery". Again the conventional wisdom, going back to Karl Popper (1968), was that particular hypotheses could be proposed and

could be falsified (definitely) or confirmed (tentatively). But the origins of those hypotheses were mysterious – there was no way of explaining how the evidence itself could generate a hypothesis.

Causal Bayes nets provide a kind of logic of inductive inference and discovery. They do so, at least, for one type of inference that is particularly important in scientific theory-formation. Many scientific hypotheses involve the causal structure of the world. Scientists infer causal structure by observing the patterns of conditional probability among events (as in statistical analysis), by examining the consequences of interventions (as in experiments) or, usually, by combining the two types of evidence. Causal Bayes nets formalize these kinds of inferences.

In causal Bayes nets, causal hypotheses are represented by directed acyclic graphs like the one below. The graphs consist of variables, representing types of events or states of the world, and directed edges (arrows) representing the direct causal relations between those variables (see figure 1). The variables can be discrete (like school grade) or continuous (like weight), they can be binary (like “having eyes” or “not having eyes”) or take a range of values (like color). Similarly, the direct causal relations can have many forms; they can be deterministic or probabilistic, generative or inhibitory, linear or non-linear. The exact specification of the nature of these relations is called the “parameterization” of the graph.

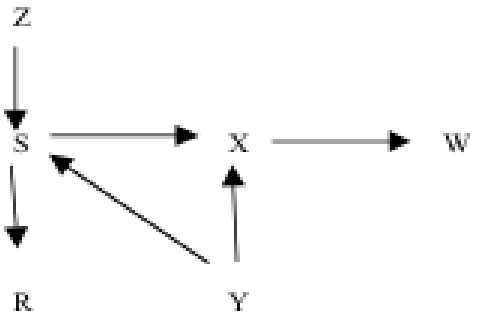


Figure 1: A causal Bayes net

Causal structure and conditional probabilities. The Bayes net formalism makes systematic connections between the causal hypotheses that are represented by the graphs and particular patterns of evidence. The structure of a causal graph constrains the conditional probabilities among the variables in that graph, no matter what the variables are or what the parameterization of the graph is. In particular, it constrains the conditional independencies among those variables. Given a particular causal structure, only some patterns of conditional independence will occur among the variables.

So the structure of the causal graph puts some very general constraints on the patterns of probability among the variables. If we make further assumptions about the parameterization of the graph, that is, about the particular nature of the causal relations among the variables, we can constrain the kinds of inferences we make still further. For

example, if we assume that each cause independently has a certain power to bring about an effect, and that that power leads to a certain likelihood of the effect given the cause, we can further constrain the patterns of conditional probability among causes and effects. This is a common assumption in studies of human causal learning.

To illustrate how this works consider a simple causal problem that is far too common for academics who attend many learned conferences. Suppose that I notice that I often can't sleep when I've been to a party and drunk lots of wine. Partying (P) and insomnia (I) covary, and so do wine (W) and insomnia (I). There are at least two possibilities about the relations among these variables which I can represent by two simple causal graphs 1) a chain P → W→ I, or 2) a common cause structure I ← P → W. Maybe parties lead me to drink and wine keeps me up; maybe parties both keep me up and lead me to drink. The covariation among the variables by itself is consistent with both these structures.

You can discriminate between these two graphs by looking at the patterns of conditional probability among the three variables. Suppose you keep track of all the times you drink and party and examine the effects on your insomnia. If graph 1 is correct, then you should observe that you are more likely to have insomnia when you drink wine, whether or not you party. If, instead graph 2 is correct you will observe that, regardless of how much or how little wine you drink, you are only more likely to have insomnia when you go to a party.

More formally, if graph #1 is right, and there is a causal chain that goes from parties to wine to insomnia, then $I \perp P \mid W$ – the probability of insomnia occurring is independent (in probability) of the probability of party-going occurring conditional on the occurrence of wine-drinking If graph #2 is right, and parties are a common cause of wine and insomnia, then $I \perp W \mid P$ – the probability of wine-drinking occurring is independent (in probability) of the probability of insomnia occurring conditional on the occurrence of party-going.

The philosopher of science Hans Reichenbach (1956) long ago pointed out these consistent relations between conditional independence and causal structure and talked about them in terms of "screening-off". When there is a chain going from partying to wine to insomnia, the wine "screens off" insomnia from the influence of partying, when partying directly causes both wine and insomnia, wine does not screen-off insomnia from partying – partying leads to insomnia directly. But partying does "screen off" insomnia from the effects of wine. Bayes nets generalize this "screening-off" principle to all acyclic causal graphs.

Thus if we know the structure of the graph, and know the values of some of the variables in the graph, we can make consistent predictions about the conditional probability of other variables. In fact, the first applications of Bayes nets involved predicting conditional probabilities (Pearl 1988). Many real life inferences involve complex combinations of conditional probabilities among variables – consider a medical expert, for example, trying to predict one set of symptoms from another set. Applying Bayes'

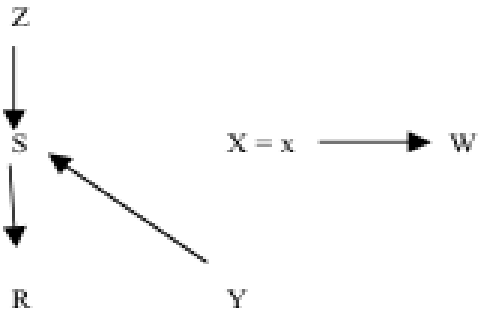
rule also requires that you track conditional probabilities. But trying to predict all the combinations of conditional probabilities rapidly becomes an exponentially complicated problem. Computer scientists were trying to find a tractable way to calculate these conditional probabilities, and discovered that representing the variables in a directed graph allowed them to do this. The graph allowed computer scientists to "read off" quite complicated patterns of conditional dependence among variables. The first applications of Bayes nets treated the graphs as calculation devices -- summaries of the conditional probabilities among events.

Bayes nets and interventions. Why think of these graphs as representations of causal relations among variables, rather than simply thinking of them as a convenient way to represent the probabilities of variables? The earlier Bayes net iterations were confined to techniques for predicting some probabilities from others. However, the development of causal Bayes net algorithms also allows us to determine what will happen when we intervene from outside to change the value of a particular variable. Here is where the interventionist account of causation comes in. We saw that on the interventionist account of causation when two variables are genuinely related in a causal way then, holding other variables constant, intervening to change one variable should change the other.

Predictions about probabilities may be quite different from predictions about interventions. For example, in a common cause structure like 2 above, we will indeed be able to predict something about the value of insomnia from the value of wine. If that structure is the correct one, knowing that someone drank wine will indeed make you

more likely to predict that they will have insomnia (since drinking wine is correlated with partying, which leads to insomnia). But intervening on their wine-drinking, forbidding them from drinking, for example, will have no effect on their insomnia. Only intervening on partying will do that.

In causal Bayes nets, interventions systematically alter the nature of the graph they intervene on, and these systematic alterations follow directly from the formalism itself. In particular, when an external intervention fixes the value of a variable it also eliminates the causal influence of other variables on that variable. If I simply decide to stop drinking wine, my intervention alone will determine the value of wine-drinking; partying will no longer have any effect. This can be represented by replacing the original graph with an altered graph in which arrows directed into the intervened upon variable are eliminated (Judea Pearl vividly refers to this process as graph surgery (Pearl, 2000)). The conditional dependencies among the variables after the intervention can be read off from this altered graph. Figure 2 for example, shows the same graph as figure 1 but now with an intervention on the variable X that sets it to a particular variable. If X is the level of drug in the patient's blood, for example, the intervention has determined that the drug will be there at particular level. We can then "read off" the graph and see what the consequences of the intervention will be (See Fig 2)



Suppose, for example, I want to know what I can do to prevent my insomnia. Should I sit in my room alone, but continue to drink when I want to or go to parties just the same but stick to Perrier? I can calculate the effects of such interventions on each of the causal structures, using "graph surgery" and predict the results. I will get different results from these interventions depending on the true causal structure (solitary drinking will lead to insomnia, and sober partying won't for graph 1, sober partying will lead to insomnia and solitary drinking won't for graph 2.)

See fig 3

Intervene on partying

P → W→ I Intervene on partying leads to P → W→ I Intervene on wine leads to W→ I

I ← P → W Intervene on partying leads to I ← P → W Intervene on wine leads to I ← P

Exactly the same inferential apparatus can be used to generate counterfactual predictions. Suppose I want to ask what would have happened, had things been otherwise. If I had refrained from wine at all those conferences would my life, or at least my insomnia, have been better? Graph surgery will answer this question too. Just as in an intervention a counterfactual "fixes" the value of certain variables and allows you to infer the consequences.

A central aspect of causal Bayes nets, indeed the thing that makes them causal, is that they allow us to freely go back and forth from evidence about observed probabilities to inferences about interventions and vice-versa. The apparatus of causal Bayes nets, then, allows us to take a particular causal structure and accurately predict the conditional probabilities of events, and also the consequences of interventions on those events, from that structure.

Bayes nets and learning

We just saw that knowing the causal structure let's us make the right predictions about interventions and probabilities. We can also use this fact to learn causal structure from the evidence of interventions and probabilities.

Lets go back to the wine-insomnia example. You could distinguish between these graphs either by intervention or observation. You could for instance, hold partying constant (always partying or never partying) and vary whether or not you drunk wine; or you could hold drinking constant (always drinking or never drinking) and vary whether or not you partied. In either case, you could observe the effect on your sleep. If drinking affects your sleep when partying is held constant, but partying has no effect on your sleep when drinking is held constant, you could conclude that graph 1 is correct. Such reasoning underlies the logic of experimental design in science.

You could also, however, simply observe the relative frequencies of the three events. If you notice that you are more likely to have insomnia when you drink wine, whether or not you party, you can infer that graph 1 is correct. If you observe that, regardless of how much or how little wine you drink, you are only more likely to have insomnia when you go to a party, you will opt instead for graph 2. These inferences reflect the logic of correlational statistics in science. In effect, what you did was to "partial out" the effects of partying on the wine/insomnia correlation, and draw a causal conclusion as a result.

Causal Bayes net representations and learning algorithms allow learners to accurately predict patterns of evidence from causal structure and to accurately learn causal structure from patterns of evidence. They constitute a kind of inductive causal logic, and a logic of causal discovery. It is possible to prove that only certain patterns of evidence will follow from particular causal structures, just as only certain conclusions follow from particular logical premises, given the axioms of logic.

It is not only theoretically possible to infer complex causal structure from patterns of conditional probability and intervention. It can actually be done. (Glymour & Cooper, 1999; Spirtes et al., 2000). Computationally tractable learning algorithms have been designed to accomplish this task and have been extensively applied in a range of disciplines (e.g., Ramsey et al. 2002; Shipley 2000). In some cases, it is also possible to accurately infer the existence of new unobserved variables that are common causes of the observed variables (Silva et al., 2003; Spirtes, Christopher, & Richardson, 2003).

Causal Bayes nets are particularly well suited to Bayesian learning techniques. (Heckerman, Meek, & Cooper 1999; Tenenbaum & Griffiths, 2003; Griffiths & Tenenbaum, 2007). Bayesian networks are designed precisely to generate probability patterns among the evidence. This means that we can use Bayesian methods to determine the probability of any particular graph given a particular pattern of contingencies among variables, or given the outcome of some set of controlled experiments.

This last point is particularly interesting. Causal Bayes nets give us an elegant and clean way of representing the outcomes of experiments. Eberhardt has systematically explored how we can use experiments to infer the causal structure of a graph from data (Eberhardt & Scheines, 2007). As we might expect, experiments turn out to be a particularly powerful tool for causal inference. If you can intervene systematically to alter the structure of some causal system you can rapidly and accurately learn how it works.

Hierarachical Bayes nets

Bayesnets are excellent representations of particular causal structures, even complex causal structures. However, often in cognitive development children are not

learning particular causal structures so much as learning abstract generalizations about causal structure. In addition to learning, for example, that my desire for milk makes me go to the fridge in search of milk, children may develop a broader generalization that desires and beliefs are the kinds of variables that lead to actions. Or they may conclude that psychological causation, in general, tends to be more probabilistic than physical causation – desires often but not always make me act, whereas a rolling ball always launches the ball it collides with.

In theories, scientific or intuitive, these broader generalizations are particularly important. In the philosophy of science, we talk about "framework theories" or even "paradigms" to capture these higher-order generalizations about particular causal structures. In cognitive development, the literature on domain-specificity suggests a similarly high-level set of generalizations. Children seem to make different assumptions about the kinds of variables and the types of causal structure to be found in psychology versus biology versus physics for example. Again this has sometimes been seen as support for the nativist idea that these high-level kinds of structure must be there innately.

Griffiths and Tenenbaum (2007) have developed ways of representing, inferring and learning higher-order generalizations about causal structure as well as the details of particular causal structures. The idea is that we can have "meta Bayes nets", representations of the structure among Bayes nets that can themselves be learned by Bayesian methods. For example, if we observe that Bayes nets that involve psychological variables also often involve indeterministic causal links, we might develop the higher-order generalization that psychological relations are likely to be indeterminate.

In turn when we evaluate the next piece of evidence about causal structure we may weigh it differently in the light of this generalization. If we see that a system seems to behave probabilistically (when we deal with a computer, for example) we may try to explain it in a psychological rather than physical way. Conversely if we know that we are dealing with a psychological system, a person, for example, we may predict that their behavior will not be perfectly consistent, even if we aren't sure just what their specific desires may be.

In fact, in a very striking set of results they call "the blessing of abstraction" Goodman et al. (2011) have shown that it is often easier to learn causal structure at several levels at once than it is to simply infer particular causal structure. Again the conventional wisdom in psychology has been that learning at a lower level of generalization and abstraction – more concrete learning – must precede learning higher-order or more abstract generalizations. This idea has been presupposed both by empiricists—who argue that therefore there are no higher-order generalizations, and by nativists who argue that such generalizations must be innate.

In fact, it is possible to show that, at least normatively, learning at several levels at once actually proceeds more quickly and efficiently than moving from the concrete to the abstract. In principle, children may be learning both specific facts about my desires, and generalizations about all desires, from the same evidence and at the same time.

Empirical Work on Bayesnets and Bayesian reasoning in children

Over the past ten years a large body of work has explored whether children might have Bayes net-like representations of causal structure, and whether they can learn causal

structure from evidence in the way that the formalism suggests. Investigators have explored whether children can learn causal structure from patterns of probability, and from the outcomes of intervention – their own or those of others. They have also investigated whether children use that knowledge to produce new interventions and experiments in ways that go beyond simple associationism. We can also ask whether children will integrate their prior knowledge with new evidence in a Bayesian way, and whether they will go beyond learning particular causal structure to learning more abstract causal schemas. The quick answer to all these questions is yes, at least, for four-year-olds.

The basic methodology of all these experiments has been similar. Obviously, it is not possible to explicitly ask very young children about topics like conditional probability. Indeed, even adults have a great deal of difficulty with explicit probabilistic reasoning. The usual method in the literature on intuitive theories has been to try to discover a typical child's knowledge of particular causal domains – such as psychology, biology and physics – at a particular age. But if we want to understand the fundamental mechanisms of causal learning we need to give children causal problems that they haven't already solved. So the methodology has been to give children controlled evidence about new causal systems, and to see what kinds of causal conclusions they draw.

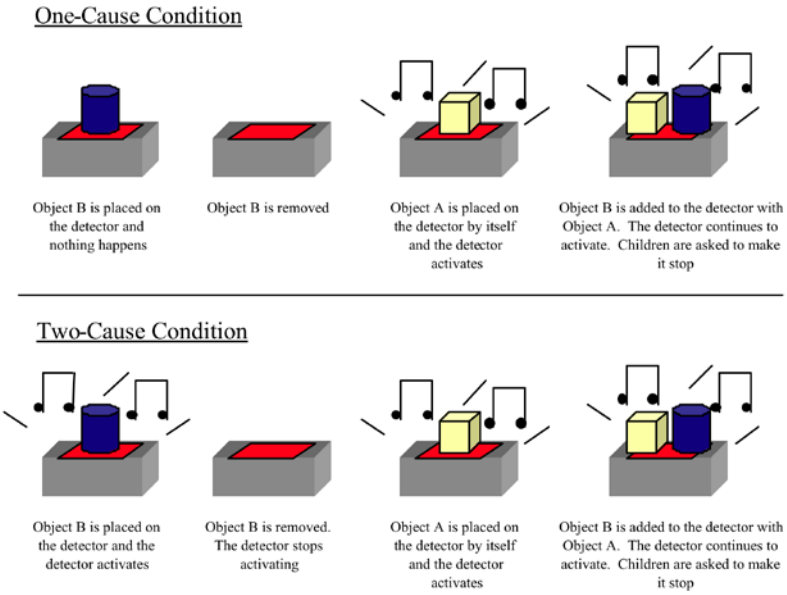
Learning Causality from Probability

We already know that children can be both classically and operantly conditioned. We also know that even infants can detect complex statistical patterns, in fact, this work on statistical learning has been one of the most important areas of developmental research in the past fifteen years (Gomez, 2002; Kirkham, Slemmer, & Johnson, 2002; Saffran,

Aslin, Newport, 1996). The interventionist approach to causation, however, suggests that children might be able to go beyond either learning the immediate consequences of their actions or associating correlated events, or even detecting statistical patterns. If children can genuinely learn new causal structure from patterns of probability they should be able to use that structure to design new interventions on the world.

In a first set of experiments Gopnik et al (2001) showed just that. Children saw a "blicket detector" -- a machine that went off when some objects but not others were placed on it. Then they were asked to design an intervention to make the machine go or turn it off. This required a kind of "screening off" reasoning. 2, 3 and 4-year-olds could use the pattern of conditional dependence between the blocks and the machine's activation to infer the causal structure of the machine, and figure out how to make it go or stop. For example, children would discriminate between the different patterns of conditional dependence illustrated in Fig 4, removing one block in the one-cause condition but both blocks in the two-cause condition. Sobel et al (2004) found that children would also make these inferences for more complex patterns of probability, particularly "backward blocking", a kind of causal inference that is not easily accommodated by associationistic accounts. Sobel & Kirkham (2006, 2007) found that children as young as 18-months-olds also showed similar capacities. Using eye tracking they also found that even 9 month-old-infants were sensitive to "screening-off" patterns, and would use them to make appropriate causal predictions – though it was more difficult to determine whether babies this young would use this information to design interventions

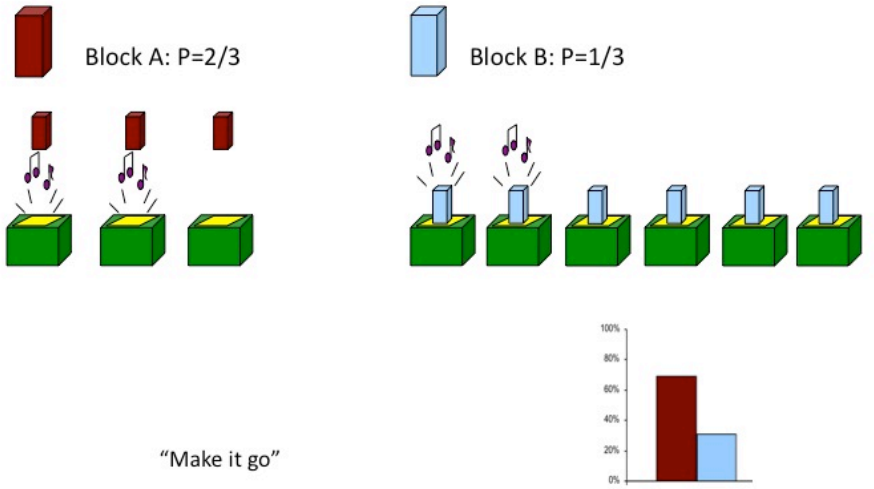
Figure 12: Procedure used in Gopnik et al. (2001), Experiment 3



Children can also infer more complicated kinds of causal structure. Schulz, Kushnir & Gopnik. (2007) showed that children could use a combination of interventions and probability to infer the direction of a causal relation, that is whether A caused B or vice-versa. And Schulz, Gopnik, & Glymour (2007), showed that four-year-old children could use this kind of evidence to infer more complex causal structures involving three variables. They could, for example, distinguish between a causal chain, where A causes B causes C and a common cause structure where A and B cause C. Finally, Kushnir & Gopnik (2005) showed that four-year-old children could use probabilistic strength to infer causal strength – they thought that a block that set off the machine 2 of 3 times was

more effective than one that worked only 2 of 6 times. They did so even when this information conflicted with information about contact.

Figure 3 Kushnir and Gopnik 2004



Children not only use probability to infer observed causal relations but they also use it to infer the existence of unobserved causes - hidden "theoretical entities". Schulz and Somerville (2006) found that when children saw an indeterministic machine – that is a machine that went off only 2 of 6 times, they inferred that some hidden variable was responsible for the failures. Similarly, Gopnik et al (2004) found that when there was no coherent way to explain a causal pattern given the observed variables, children would look for unobserved variables.

Comment [AG1]: Gopnik and Schulz or Gopnik et al 2004?

Learning across domains

Many of these experiments involved physical causation of the sort that we would see in a machine. However, children also extended these causal learning techniques to other domains. Three and four year old children can use "screening off" to infer mental states such as desires, emotions and traits, and biological relations such as sneezing (Schulz & Gopnik 2004, Seiver et al 2009). In a striking set of experiments Kushnir et al. (2010), found that children as young as 20 months old would use statistical Bayesian reasoning to infer the desires of another person.

Xu & Garcia (2008) originally demonstrated that infants as young as 9 months old were sensitive to patterns of statistical sampling. Infants looked longer when a sample of mostly red balls was taken from a box of mostly white balls, than when a sample of mostly white balls was extracted. Kushnir et al (2010) found that 20-month-olds interpreted this non–random sample causally – they thought it must be the result of an underlying preference for red balls over white ones. The experimenter took mostly green frogs from a box of mostly yellow ducks and extended her hand. The children could give her either a frog or a duck. When she took mostly frogs from a box that held mostly ducks, they gave her a frog. When she took mostly frogs from a box that held mostly frogs, children were equally likely to give her the frog or the duck.

Intervention and Experimentation

In these experiments children were exposed to patterns of probability that were the outcome of interventions- an experimenter would perform a series of action on objects and some set of consequences would ensue. A number of experiments also suggest that, at least by age 4, children can distinguish the information to be gained from interventions and probabilities and can combine the two sources of information

effectively (Gopnik et al. 2004). In fact, there is even surprising evidence that rats can make this distinction in some cases (Blaisdell et al, 2006). Moreover, four-year-olds discriminate between their own interventions and those of others and recognize that confounded interventions may not be causally informative (Kushnir & Gopnik, 2005).

Comment [AG2]: Gopnik and Schulz or Gopnik et al 2004?

Schulz has undertaken a particularly interesting research program exploring the ways in which children's spontaneous exploratory play can be thought of as a kind of intuitive experimentation and can help children discover causal structure (Schulz & Bonawitz, 2007; Schulz, Gopnik & Glymour 2007; Schulz, Standing, & Bonawitz, 2008). Schulz & Bonawitz (2007) assessed how preschool children explored a new "jack-in-the-box" type of toy that had two levers and two effects (a duck and/or a puppet could pop up). Crucially, they compared two conditions, one where the causal structure of the toy was ambiguous and one where it was clear. In the *confounded* demonstration, an adult and the child pushed both levers simultaneously and both effects appeared. With this demonstration it was completely unclear if one lever produced the duck, and the other produced the puppet, if one lever produced both effects, if one lever could be pushed independent of the other, if one effect could appear independent of the other, etc. In the *unconfounded* condition the adult pushed one lever and it systematically produced a single effect, and then the child pushed the other which systematically produced the other effect.

Thus in the unconfounded condition the causal structure of the toy was clear and in the confounded condition it was ambiguous. This "old" toy and a new one (a different, simpler, single-lever toy) were then placed in front of the child. The child was left alone, free to play with either toy. If children's play just depends on novelty, they should play

with the “new” toy. But if their play is driven by a desire to understand causal structure, then, in the confounded condition, they should be especially likely to explore the “old” toy, because in that condition the old toy’s causal structure is unclear. Indeed, 3- and 4-year-old children systematically explored the old toy rather than the new one in the confounded condition but not the unconfounded one. Moreover, after they had finished exploring the toy children in the confounded condition showed that they had used the results of their exploration to figure out how the toy worked.

When children were simply given a causally puzzling machine to play with, they spontaneously produced interventions on that system – experiments -- that revealed the causal structure of that machine, and they did so in intelligent ways. This research jibes nicely with Eberhardt’s theoretical research showing the formal benefits of using experimentation for causal inference (Eberhardt & Scheines, 2007). It also shows how research inspired by probabilistic models can shed light on classical developmental issues. For many years developmentalists and educators have felt that exploratory play must help children learn but actual demonstrations of this have been rare. Schulz et al. show systematically that this is true for causal inference.

Limitations on causal inference

So far everything I’ve reported has gone to show that even very young children are far more sophisticated causal learners than we would have thought. There is, however, at least one interesting set of findings suggesting a surprising limitation in children’s causal inference. In the experiments I’ve described so far children see

probability patterns that are the result of human actions, though not their own actions. They see an experimenter put a block on the machine, several times, for example, and then observe the pattern of probability of the outcomes.

Surprisingly toddlers, and perhaps also non-human animals, seem to have a great deal of difficulty making causal inferences from "pure" correlations- correlations that are not the result of human actions. Bonawitz et al (2010), showed two-year-olds simple correlations between two spontaneous events that were not the outcome of actions. A car would collide with a box, the box would light up and then a toy plane a few inches away would spin, without any human action being involved. The children then had the opportunity to make one box collide with the other, in order to make the plane spin. Although the children would move the boxes happily when they were asked to, they did not move the box in order to make the plane go. They did do this when the system was described in explicitly causal terms ("Look the box lit up and it made the plane spin!") and when the box and the plane were spatially contiguous. Meltzoff et al (2007) found a similar pattern of results with different stimuli and also found that 24-month-olds would make theses inferences correctly when a human agent brought about the pattern of results.

Integrating prior knowledge and new evidence

There are also several studies showing that children take prior knowledge into account in a Bayesian way when they are making causal inferences, but that, also in a Bayesian way, they will overturn that prior knowledge given enough evidence. In the "backwards blocking" experiment Sobel et al (2004) showed that children would take the baseline frequency of blickets into account when they made new inferences – they made

different inferences when they were told beforehand that blickets were rare or common. Moreover, Schulz & Gopnik (2004) showed that children at least as young as 4 would use conditional probability information to make cross-domain inferences. They inferred, for example, that anxiety could cause an illness or that talking to a machine could make it go. These cross–domain inferences are a good example of an initially low probability hypotheses that may be confirmed by the right pattern of evidence. Schulz, Bonawitz, and Griffiths (2007), also later showed how a gradual accumulation of evidence could shift children's inferences about these cross-domain cases.

Similarly, Kushnir and Gopnik (2005) found that children initially preferred a hypothesis about a machine that involved contact, as we might expect on Michottean grounds. They assumed that a block would have to touch the blicket detector to make it go. However, they rapidly overcame that prior assumption when they were presented with probabilistic contingency evidence that supported a hypothesis involving action at a distance. When they saw that the machine activated most frequently when an object was waved above it, rather than touching it, they concluded that contact was unnecessary.

Finally, at least two recent studies show that four-year-old children can go beyond specific kinds of causation to infer abstract generalizations in the way that hierarchical Bayes nets would propose. Schulz et al. (2008) designed an experiment in which blocks from different underlying and non-obvious categories would interact differently – two blocks could always activate each other, for example, or only some blocks would activate other ones. Children could use the data to discover the underlying categories and generalized these high-level rules to new cases. Lucas et al (2009) similarly found that children who saw a "blicket detector" the followed either a disjunctive rule (all red blocks

make it go) or a conjunctive one (you need both a red and blue block to make it go) generalized this higher-order rule to new cases.

Future Directions

So far this body of work has been remarkably productive, generating a wide variety of new questions and interesting empirical results. Many questions still remain however, both at the empirical and computational level.

The search problem

Understanding process

So far, work in the probabilistic models program has focused on whether children can behave like rational causal inference machines. Can they draw the right conclusions from the data they see? To a remarkable degree, the answer is yes. A question that remains is how children actually manage to do this. As I said earlier, the most serious drawback of the probabilistic model approach is the search problem. It is intractable for any learner, human or artificial, to systematically search through the entire space of hypotheses. Where do the hypotheses come from? How do children move from one hypothesis to another? And how do they restrict their search space?

A common solution to this problem in computer science is some form of hypothesis sampling. The picture is that a system begins by testing a few hypotheses against the evidence. This sample is random, but respects the probability distribution of the hypotheses. So you would sample more high probability hypotheses but also include a few low probability ones just in case. Then you update the probability of these sampled hypotheses given the evidence. Considering the evidence will also change the probability of all the hypotheses, some will become more likely and some less so. So another sample,

taken after the updating has taken place, will come up with a different set of hypotheses, and you can test those new hypotheses against new evidence, and so on. Children often make apparently random multiple guesses when they try to solve a problem and this led previous investigators to dismiss children as irrational. Instead, however, as Siegler (1989) has suggested these children might instead be sampling from a range of hypotheses, and trying these hypotheses against the data.

Alternatively one might have constructive mechanisms that actually build up structures like causal graphs from the evidence directly. Processes of analogy might also help children to import structure from one domain to another – if a particular causal structure is well confirmed in the physical world for example, we might try applying it to a psychological problem.

Social Cognition and Knowledge

A rather different way of constraining the search space, and deciding which variables to attend to, is to use external cues and particularly social cues as a way to guide causal inference. There has been a substantial amount of recent work suggesting that such cues might play a very important role in guiding causal inference and this is also a particularly exciting direction for future work. In particular, the results of Bonawitz et al (2010), and Meltzoff et al. (under review) suggest that, especially for young infants and toddlers, the outcomes of human actions might be a particularly strong cue to causal significance. The substantial literature on imitation in infancy shows that human infants are particularly adept at learning from the actions of others, in general, and this might be an especially important way to learn about causal structure (see eg Woodward, 1998). By restricting themselves to these events, rather than attending to all the patterns of

correlation in the world around them, children could restrict their inferences to a few relevant dimensions.

Causal language might play a similar role. A long tradition, in general, and several recent studies, in particular, suggest a close link between language and causal inference. Causal language, like adult actions, could lead children to only pay attention to some patterns of probability and not others. If children only pay attention to correlations that result from actions or that are labeled as causal that would be a substantial way of constraining the search space. It would also ensure that children learned the causal structure that was most relevant for their particular culture.

There is also a very interesting line of work exploring the interaction between causal inference and "intuitive pedagogy". Shafto and colleagues (2010) have shown that a rational system should make different inferences when it assumes that the evidence it sees comes from a teacher, rather than simply being generated by the world itself. The assumption is that evidence that is provided by a teacher will be carefully selected rather than randomly generated. Children may make similar assumptions – after all much of the evidence they see is coming from knowledgeable teachers -- and again this would constrain their hypothesis space in interesting and helpful ways. Some recent studies (Buchsbaum et al, 2009; in press) suggest that children may indeed do this.

Integrating Mechanism, Explanation and Intuitive Theories

The causal Bayes net framework provides a coherent way of thinking about agency, counterfactuals, probability and association. However, this work still needs to be integrated with the other relevant bodies of research on causation and causal development. Where does mechanism fit in, for example? How do these inferences mesh

with the perceptual causality work in the Michottean tradition? There are several ways of thinking about this relationship. It is possible that the two types of cognition are profoundly different. In particular, perceptual causality may be thought of as a kind of module – an encapsulated part of the visual system rather than a part of more conceptual causal understanding (Scholl, 2000). In support of this idea there is some evidence that quite different brain areas are involved in Michottean predictions and "blicket detector" probabilistic causal inferences (Blakemore et al, 2001; Roser et al. 2005). Not surprisingly high-level visual areas and right hemisphere processes appear to correlate with perceptual causality while areas like left dorsolateral prefrontal cortex, areas that are more involved in high-level reasoning, appear to be activated in contingency learning.

It is also possible that the Michottean effects are simply the result of patterns of covariation that are particularly ubiquitous in the visual world. As a result these patterns might be learned at a very early age, and would be very highly confirmed and so difficult to change by adulthood.

On the other hand the strength of our causal intuitions about mechanism and space is striking, and appears to go beyond merely modular perceptual processes or histories of contingency. It may be that we assume that causal influence, defined in terms of intervention, will map on to spatial structure, even if we aren't sure how this mapping will work in detail. It is also possible that physical mechanisms have special features that make them particularly robust and reliable causal systems for purposes of intervention (Woodward, 2002). We might be especially likely to detect the causal relations in these systems, either in evolutionary time or ontogenetically.

A second set of questions concerns the relation between explanation and causation. Traditionally, explanation has been seen as a rather sophisticated language-intensive activity. But work by Schult and Welman (1997) and Hickling and Wellman (2001) shows that children as young as two will vigorously and regularly produce explanations. Le Gare (2010, in press) has shown that children are most likely to produce explanations when they see anomalous evidence and, most significantly, that explanations improve children's causal learning. Explanation and causation are obviously intimately connected, explanations provide an account of the causal history of an event. But exactly how explanations work and why they appear to generate so powerful an advantage is still mysterious.

A final set of questions involves the integration of causal inference and the large existing body of work on intuitive theories. We know from the probabilistic models work that children have powerful causal learning abilities. We know from work on intuitive theories that children have extensive causal knowledge of the physical, biological and psychological world, and that this knowledge changes with development. A plausible conclusion is that these learning abilities are responsible for those changes. There are a few studies showing that causal inference could lead to such theory changes as the development of a representational theory of mind at 4 (Goodman et al. 2006) or of a trait theory of personality at 6 or 7 (Seiver et al. 2009). But much more remains to be done to demonstrate that this is so.

Conclusion

The riddle of causality will no doubt continue to puzzle developmentalists, as it has long puzzled philosophers, and indeed as it continues to puzzle children themselves.

There is, as always in developmental psychology, a deep and satisfying affinity between we scientists and our subjects. As they try, with exploratory curiosity and explanatory joy, to work out the causal structure of our world, we will, with much the same mix of emotions, continue to work out the causal structure of their world.

Questions for further research

1. What is the relationship between mechanistic and probabilistic conceptions of causality?

2. What are the actual algorithms used in causal learning? How are they implemented in the brain?

3. How do social factors such as implicit pedagogy influence causal learning?

4. Is the idea of causality innate or could it itself be constructed from experience?

5. To what extent is causal knowledge and learning similar or different in different domains such as the physical, biological and psychological realms?

Take-away Points

1. Although there is general agreement that causal knowledge is important there are many different conceptions of such knowledge.
2. Piaget and others suggested that children's causal knowledge is rooted in agency and this is implicit in operant conditioning and trial and error theories of learning.
3. Michotte, Leslie, Gelman and Schulz suggest that it involves particular mechanical or physical events involving contact, launching and the transmission of force, and may be innate.
4. Classical conditioning and connectionist accounts suggest that it is simply learned through correlation and association.
5. Probabilistic model accounts suggest that children's causal knowledge is related to intervention and counterfactual reasoning.
6. Probabilistic models, such as causal Bayes nets, represent causal hypotheses as structured directed graphs which generate evidence.
7. Causal graphs generate special inferences about interventions and counterfactuals through graph surgery.

8. Causal graphs can be learned from patterns of evidence using Bayesian techniques.
9. Preschool children, and perhaps infants, infer causal structure from correlations and the outcome of interventions, their own and others.
10. Children's causal learning can be explained, at least in part, by probabilistic models.

References

Baillargeon, R. (2008). Innate ideas revisited: For a principle of persistence in infants' physical reasoning. *Perspectives on Psychological Science*, *3*(1), 2-13.

Blaisdell, A. P., Sawa, K., Leising, K. J., & Waldmann, M. R. (2006). Causal reasoning in rats. *Science*, *311*(5763), 1020-2.

Blakemore, S.J., Fonlupt, P., Pachot-Clouard, M., Darmon, C., Boyer, P., Meltzoff, Andrew N., et al. (2001). How the Brain Perceives Causality: An Event-Related fMRI Study. *Neuroreport*, *12*(17), 3741-3746.

Blass, E., Ganchroa, J., & Steiner, J. (1984). Classical conditioning in newborn humans 2–48 hours of age. *Infant Behavior and Development*, *7*(2), 223-235.

Bonawitz, E. B., Ferranti, D., Saxe, R., Gopnik, A., Meltzoff, A. N., Woodward, J., et al. (2010). Just do it? Investigating the gap between prediction and action in toddlers' causal inferences. *Cognition*, *115*(1), 104-17.

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.

Buschbaum, D., Griffiths, T. L., & Gopnik, A. (In Press). Children's Imitation of Causal Action Sequences is Influenced by Statistical and Pedagogical Evidence. *Cognition.*

Buschbaum, D., Griffiths, T. L., Gopnik, A., & Baldwin, D. (2009). Learning from Actions and their Consequences: Inferring Causal Variables from a Continuous Sequence of Human Action. In N. A. Taatgen & H. Van Rijn (Eds.), *Proceedings of*

the 31st annual conference of the Cognitive Science Society (p. 2493–2498). Austin, TX: Cognitive Science Society.

Chomsky, N. (2006). *Language and Mind* (3rd ed.,). Cambridge: Cambridge University Press.

Danks, D. (2003). Equilibria of the Rescorla–Wagner model. *Journal of Mathematical Psychology*, *47*(2), 109-121.

Dowe, P. (2007). *Physical Causation*. Cambridge; New York: Cambridge University Press.

Eberhardt, E., & Scheines, R. (2007). Interventions and Causal Inference. *Philosophy of Science*, 981-995.

Elliott M. Blass. (1984). Classical conditioning in newborn humans 2–48 hours of age. *Infant Behavior and Development*, *7*(2), 223-235.

Flavell, J. H. (1999). Cognitive development: Children's knowledge about the mind. *Annual review of psychology*, *50*, 21-45.

Gelman, S. A. (2003). *The Essential Child: Origins of Essentialism in Everyday Thought (Oxford Series in Cognitive Development)* (1st ed., p. 392). New York: Oxford University Press.

Gelman, S., & Raman, L. (2002). Folk Biology as a Window Into Cognitive Development. *Human Development*, *45*(1), 61-68.

Glymour, C. N. (2001). *The Mind's Arrows: Bayes Nets and Graphical Causal Models in Psychology* (1st ed., p. 222). MIT Press.

Glymour, C., & Cooper, G. (Eds.). (1999). *Computation, Causation, and Discovery*. Menlo Park, CA: AAAI Press/ MIT Press.

Gomez, R. L. (2002). Variability and Detection of Invariant Structure. *Psychological Science*, *13*(5), 431-436.

Goodman, N. D., Ullman, T. D., & Tenenbaum, J. B. (2011). Learning a Theory of Causality. *Psychological Review*, *118*(1), 110-9.

N. D. Goodman, C. L. Baker, E. B. Bonawitz, V. K. Mansinghka, A. Gopnik, H. Wellman, L. Schulz, and

J. B. Tenenbaum (2006). Intuitive theories of mind: A rational approach to false belief. Proceedings of

the Twenty-Eighth Annual Conference of the Cognitive Science Society.

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), 3-32.

Gopnik, A., & Melzoff, A. (1997). *Words Thoughts and Theories*. Cambridge, MA: MIT Press.

Gopnik, A., & Schulz, L. (2004). Mechanisms of Theory Formation in Young Children. *Trends in Cognitive Sciences*, *8*(8), 371-7.

Gopnik, A., Sobel, D., Schulz, L., & Glymour, C. (2001). Causal learning mechanisms in very young children: Two, three, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology*, *37*(5), 620-629.

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in Cognitive Sciences*, *14*(8), 357-64.

Griffiths, T. L., & Tenenbaum, J. B. (2007). Two Proposals for Causal Grammar. In A. Gopnik & L. Schulz (Eds.), *Causal Learning Psychology, Philosophy, and Computation* (p. 323). New York: Oxford University Press.

Han. F. & Zhu, S. C. (2005). Bottom-up/top-down image parsing by attribute graph grammar. In Proceedings of the Tenth IEEE International Conference on Computer Vision, pp. 1778–1785.

Harris, P., German T. & Mills P. (1996). Children's use of counterfactual thinking in causal reasoning. Cognition, 61(3), 233-259.

Heckerman, D., Meek, C. & Cooper, G. (1999). A bayesian approach to causal discovery. In Glymour, C., & Cooper, G. (Eds.). (1999). *Computation, Causation, and Discovery*. Menlo Park, CA: AAAI Press/ MIT Press.

Hitchcock, C. (2007). Prevention, preemption, and the principle of sufficient reason. *Philosophical Review*, *116*(4), 495-532.

Hickling, A. K., & Wellman, H. M. (2001). The emergence of children's causal explanations and theories: Evidence from everyday conversation. *Developmental psychology*, *37*(5), 668-683. American Psychological Association.

Hume, D. (1740). *An Abstract of A Treatise of Human Nature*.

Kemp, C. & Tenenbaum, J. B (2009). Structured statistical models of inductive reasoning. Psychological Review, 116:20–58, 2009.

Kirkham, N., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35-B42.

Comment [AW3]: Does this need publisher info? Cant find.

Kushnir, T., & Gopnik, A. (2005). Young Children Infer Causal Strength From Probabilities and Interventions. *Psychological Science*, *16*(9), 678-683.

Kushnir, T., Xu, F., & Wellman, H. M. (2010). Young Children Use Statistical Sampling to Infer the Preferences of Other People. *Psychological Science*, *21*(8), 1134-40.

Legare, C.H. (in press). Exploring explanation: Explaining inconsistent information guides hypothesis-testing behavior in young children. *Child Development*.

Legare, C.H., Gelman, S.A., & Wellman, H.M. (2010). Inconsistency with prior knowledge triggers children's causal explanatory reasoning. *Child Development, 81*, 929-944.

Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition*, *25*(3), 265-88.

Leslie, A. M. (1994). ToMM, TOBY, and Agency: Core architecture and domain specificity. In L. A. Hirschfeld & Susan A. Gelman (Eds.), *Mapping the mind: domain specificity in cognition and culture* (pp. 119-148). Cambridge: Cambridge University Press.

Lewis, D. K. (1973). *Counterfactuals*. Maldden, MA: Basil Blackwell Ltd.

Lucas, C. G., Gopnik, A., & Griffiths, T. L. (2009). Developmental differences in learning the forms of causal relationships. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*.

Mast, V. K., Fagen, J. W., Rovee-Collier, C. K., & Sullivan, M. W. (1980). Immediate and Long-Term Memory for Reinforcement Context: The Development of Learned Expectancies in Early Infancy. *Child Development*, *51*(3), 700 - 707.

Meltzoff, A. N., & Blumenthal, E. J. (2007). Causal understanding and imitation: Effect monitoring in infants. Paper presented at Harvard University pre-SRCD meeting, Cambridge, MA.

Meltzoff, A. N., Gopnik, A., & Waisman, A. (Under review). Learning About Causes From People: Observational Causal Learning in 24-Month-Old Infants.

Menzies, P., & Price, H. (1993). Causation as a Secondary Quality. *The British Journal for the Philosophy of Science*, *44*(2), 187 - 203.

Michotte, A. E. (1962). *Causalite, permanence et realite phenomenales, eutdes de psychologie experimentale*. Louvian: Publications Universitaries .

Oakes, L. M., & Cohen, L. B. (1990). Infant Perception of a Causal Event. *Cognitive Development*, *5*(2), 193-207.

Palmer, S. E. (1999). *Vision science: Photons to phenomenology* (p. 810). Cambridge, MA: MIT Press.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (p. 552). Morgan Kaufmann.

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

Piaget, J. (1929). *The Child's Conception of the World* . London: Routledge and Kegan Paul, Ltd.

Piaget, J. (1936). *Origins of Intelligence in Children*. New York: International Universities Press.

Pinker, S. (1996). *Language learnability and language development.* Cambridge: Harvard University Press

Popper, K. R. (1968). *Conjectures and refutations: The growth of scientific knowledge*. New York: Harper & Row.

Ramsey, J., Roush, T., Gazis, P., & Glymour, C. (2002). Automated remote sensing with near-infra-red reflectance spectra: Carbonate recognition. *Data mining and knowledge discovery*, *6*, 277-293.

Reichenbach, H. (1956). *The Direction of Time* (1st ed., p. 271). Berkeley, CA: University of California Press.

Richardson, T., with P. & Christopher M. (1997). Causal inference in the presence of latent variables and selection bias. In Uncertainty in artificial intelligence: Proceedings of the eleventh conference. San Francisco, CA: Morgan Kaufmann, 499-506.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic Cognition: A Parallel Distributed Processing Approach* (1st ed., p. 427). Cambridge, MA: MIT Press.

Roser, M. E., Fugelsang, J. A., Dunbar, K. N., Corballis, P. M., & Gazzaniga, M. S. (2005). Dissociating processes supporting causal perception and causal inference in the brain. *Neuropsychology*, *19*(5), 591-602.

Rovee-Collier, C. K., & Sullivan, M. W. (1980). Organization of Infant Memory. *Journal of Experimental Psychology: Human Learning & Memory*, *6*(6), 798-807.

Rovee-Collier, C., Sullivan, M., Enright, M., Lucas, D., & Fagen, J. (1980). Reactivation of Infant Memory. *Science*, *208*(4448), 1159-1161.

Rozenblit, L., & Keila, F. (2002). The Misunderstood Limits of Folk Science: An Illusion of Explanatory Depth. *Cognitive Science*, *26*(5), 562-521.

Russell, B. (1912). On the Notion of Cause. *JSTOR: Proceedings of the Aristotelian Society, New Series*, *13*, 1-26. Blackwell Publishing on behalf of The Aristotelian Society.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science*, *274*(5294), 1926-1928.

Salmon, W. C. (1998). *Causality and explanation* (p. 434). Oxford University Press.

Schlottmann, A. (1999). Seeing it Happen and Knowing How it Works: How Children Understand the Relation Between Perceptual Causality and Underlying Mechanism. *Developmental Psychology*, *35*(5), 303-317.

Scholl, B. J., & Nakayama, K. (2002). Causal Capture: Contextual Effects on the Perception of Collision Events. *Psychological Science*, *13*(6), 493-498. SAGE Publications.

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual Causality and Animacy. *Trends in cognitive sciences*, *4*(8), 299-309.

Schult, C. A., & Wellman, H. M. (1997). Explaining human movements and actions: Children's understanding of the limit s of psychological explanation. Cognition, 62(3), 291-324.

Schulz, L. E., & Bonawitz, E. B. (2007). Serious Fun: Preschoolers Engage in More Exploratory Play When Evidence is Confounded. *Developmental Psychology*, *43*(4), 1045-50.

Schulz, L. E., Bonawitz, E. B., & Griffiths, T. L. (2007). Can being scared cause tummy aches? Naive theories, ambiguous evidence, and preschoolers' causal inferences. *Developmental psychology*, *43*(5), 1124-39.

Schulz, L. E., & Gopnik, A. (2004). Causal learning across domains. *Developmental psychology*, *40*(2), 162-76.

Schulz, L. E, Gopnik, A., & Glymour, C. (2007). Preschool Children Learn About Causal Structure From Conditional Interventions. *Developmental Science*, *10*(3), 322-32.

Schulz, L. E., Goodman, N. D., Tenenbaum, J. B., & Jenkins, A. C. (2008). Going Beyond the Evidence: Abstract Laws and Preschoolers' Response to Anomalous Data. *Cognition*, *109*(2), 211-223.

Schulz, L. E. , Kushnir, T., & Gopnik, A. (2007). Learning from doing: Intervention and causal inference in children. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, computation*. New York: Oxford University Press.

Schulz, L. E., & Sommerville, J. (2006). God Does not Play Dice: Causal Determinism and Preschoolers' Causal Inferences. *Child Development*, *77*(2), 427-42.

Schulz, L. E., Standing, H. R., & Bonawitz, E. B. (2008). Word, Thought, and Deed : The Role of Object Categories in Children's Inductive Inferences and Exploratory Play. New perspectives on the effects of action on perceptual and cognitive development. *Developmental Psychology*, *44*(5), 1266-1276.

Seiver, E., Gopnik, A., & Lucas, C. (2009). Social causal models in children: understanding the interaction of person and situation. Poster at the Society of Research in Child Development. Boston, MA.

Shafto, P., Goodman, N. D., Gerstle, B., & Ladusaw, F. (2010). Prior expectations in pedagogical situations. 32nd annual conference of the Cognitive Science Society.

Shanks, D. R. (1995). Is Human Learning Rational? *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *48*(2), 257-279.

Shanks, D. R., & Dickinson, A. (1987). Associative accounts of causality judgment. In G.H. Bower (Eds.), *Accounts of causality judgment. In G.H. Bower (Ed.)*, (pp. 229-261). New York: Academic Press.

Shipley, B. (2000). *Cause and correlation in biology*. Oxford, England : Oxford University Press.

Shultz, T. R. (1982). Rules of Causal Attribution. *Monographs of the Society for Research in Child Development*, *47*(1), 1 - 51. Blackwell Publishing on behalf of the Society for Research in Child Development.

Siegler, R. S., & Jenkins, E. A. (1989). *How Children Discover New Strategies*. New Jersey: Routledge.

Silva, R., Scheines, R., Glymour, C., & Spirtes, P. (2003). Learning measurement models for unobserved variables. *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence (AAAI-2003).* AAAI Press.

Skinner, B. F. (1938). *The Behavior of Organisms: An Experimental Analysis* (p. 457). Cambridge, Massachusetts: B.F. Skinner Foundation.

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. (2006). Blickets and babies: The development of causal reasoning in toddlers and infants. *Developmental Psychology, 42*(6), 1103.

Sobel, D. M., & Kirkham, N. Z. (2007). Bayes nets and babies: infants' developing statistical reasoning abilities and their representation of causal knowledge. *Developmental science*, *10*(3), 298-306.

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*(3), 303-333.

Spirtes, P., Christopher M. & Richardson, T. (2003). Causal inference in the presence of latent variables and selection bias. *In Uncertainty in artificial intelligence: Proceedings of the eleventh conference.* (pp. 499-506). San Francisco, CA: Morgan Kaufmann.

Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search 2nd Edition*. Cambridge: MIT Press

Kushnir, T. & Gopnik, A. (2007). Conditional probability versus spatial contiguity in causal learning: Preschoolers use new contingency evidence to overcome prior spatial assumptions. *Developmental Psychology*, *43*(1), 186-196.

Tennenbaum, J. B., & Griffiths, T.L. (2003). Theory-Cased Causal Inference. In S. Becker, S. Thrum, & K. Obermayer (Eds.), *Preceedings of the 2002 Neural Information Processing System Conference (Vol. 15)* (pp. 35-42). Cambridge, MA: MIT Press.

Watson, J. S., & Ramey, C. T. (1987). Reactions to response-contingent stimulation in early infancy. In J. Oates & S. Sheldon (Eds.), *Cognitive development in infancy*. Hove, England: Lawrence Erlbaum Associates, Inc.

Wellman, H. J. (1992). *The Child's Theory of Mind* (2nd ed.,). MIT Press.

Wellman, H. M., & Liu, D. (2004). Scaling of Theory-of-Mind Tasks. *Child Development*, *75*(2), 523-541.

Wellman, H., & Inagaki, K. (Eds.). (1997). *The Emergence of Core Domains of Thought: Children's Reasoning About Physical, Psychological, and Biological Phenomena. New Directions for Child Development, No 75*. San Francisco, CA: Jossey-Bass/Pfieffer.

White, P. A. (1995). *The Understanding of Causation and the Production of Action: From Infancy to Adulthood* (p. 320). Psychology Press.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69*(1), 1-34.

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation* (p. 413). New York: Oxford University Press.

Woodward, J. (2002). What is a mechanism? A counterfactual account. *Philosophy of Science*

Vol. 69, No. S3. pp. S366-S377

Xu, F., & Garcia, V. (2008). Intuitive Statistics by 8-month-old Infants. *Proceedings of the National Academy of Sciences of the United States of America*, 105(13), 5012-5.

Xu. F., & Tenenbaum, J. B (2007). Word learning as Bayesian inference. *Psychological Review*, 114:245–272,