Words, kinds and causal powers: A theory theory perspective on early naming and categorization.

Alison Gopnik

Dept. of Psychology

University of California at Berkeley

Thierry Nazzi

Laboratoire Cognition et Développement

CNRS – Université Paris 5

Acknowledgements: Clark Glymour and Michael Strevens made extremely helpful comments on and contributions to this paper. We are also grateful to Tamar Kushnir, Laura Schulz, and David Sobel, as well as all of the participants in our studies.

Address for correspondence: Alison Gopnik, Dept. of Psychology, University of California at Berkeley, Berkeley, CA, 94720. Gopnik@socrates.berkeley.edu

In Rakison, David H.; Oakes, Lisa M. (2003). Early category and concept development: Making sense of the blooming, buzzing confusion. (pp. 303-329). New York, NY, US: Oxford University Press. xxi, 442 pp.

Words, kinds and causal powers: A theory theory perspective on early naming and categorization.

For some twenty-five years, the prevailing theories of categorization in philosophy have invoked the idea of "kinds" (Putnam, 1975; Kripke, 1972). When we look at how adults use words to refer to categories of things we find that they only rarely categorize objects on the basis of their common properties. Instead, adults seem to categorize objects together when they believe that they belong to the same "kind"; that is, that they share some common, abstract "essence." Psychological investigations of adults have largely confirmed these philosophical intuitions, adults do seem to group objects together based on "kinds" rather than properties (Murphy & Medin, 1985, Rips, 1989).

Several investigators, particularly Gelman and her colleagues (Gelman & Markman, 1986; Gelman & Coley, 1990), have argued that children as young as two also categorize and name objects based on kinds (see also Soja, Carey, & Spelke, 1991; Mervis & Bertrand, 1994; Mandler, 1993). Other investigators, in contrast, have suggested that young children categorize and name objects on the basis of perceptual properties such as shape (Landau, Smith, & Jones, 1988; Imai & Imai et al. 1994). However, the question of whether children use names to refer to kinds has been bedeviled by the problem that the philosophical notion is not easily translatable into empirical predictions. In the philosophical literature, the concept of kinds refers to the idea that members of a category share an "essence," a common, abstract, ontological property, and that names refer to this essence (Putnam, 1975). Obviously, it is difficult to ask three-year-olds if they conceive of objects or use names in this way.

This methodological problem also raises a deeper question: why would adults or children organize the world in term of kinds? Why invoke mysterious, invisible, underlying essences when perceptual properties would do? What functional or evolutionary basis could there be for such a cognitive practice?

We will argue that, at least on many occasions, categorizing objects into "kinds" should be seen as making a claim about the common causal powers of those objects (see also Ahn et al. 2000; Strevens, 2000; Rehder, in press). On the causal powers view, we assume that something about an object causes the object's characteristic properties. This common underlying cause is responsible for the correlation among the properties of the object. This common underlying cause is also shared by other objects of the same kind and is responsible for the patterns of correlation of properties among all of those objects. We can think of the "essence" of a category as a sort of causal placeholder. To say that members of a kind share an "essence" is to say that their characteristic properties are due to a common cause, even if we may not be able to exactly identify that cause.

Different types of causal relations may be involved in different types of "kinds". For biological natural kinds such as tigers and lemons, the causal relations are likely to involve the detailed internal structure of the object, as well as to involve biological causal processes such as growth and inheritance (Gelman & Wellman, 1997). For physical or material natural kinds, they involve the material composition of the object (Soja, Carey, & Spelke, 1991). For artifacts, these relations involve intentional causal processes, in particular, the intentions of the designer of the object and the goals the object is intended to serve (Gelman & Bloom, 2000; Kemler-Nelson, 1999; Kemler-Nelson et al. 2000a, b).

One characteristic of kinds is that they support induction. An object that is a member of a kind will prove to have the same novel properties as other members of the kind (see e.g. Gelman & Markman, 1986). However, the justification for these inductive inferences has been rather mysterious. Why would we assume that an object that has the same name as another object would also prove to have the same novel properties? Assuming that the common properties of kind members are due to an underlying common cause licenses such inductions. If we now discover that members of the kind consistently have some other property in common, we can assume that that property is also caused by the object's "essence". This causal inference allows us to predict that the new property will also be present in new members of the kind.

The causal powers view predicts that prior knowledge about causal relations will influence categorization. For example, there are several studies which suggest that for adults and even older children, who already know about causal relations, common causes are more likely to be important in defining a kind category than common effects (Ahn et al., 2000; Rehder, 1999). Common DNA or common internal structure is more likely to define category membership than common color, just because DNA or internal structure is more likely to be a cause of other properties of the object than color.

But, on our view, we may also work backwards from the correlational patterns among properties to inferences about causal structure -- even if we don't know exactly which causal mechanisms are involved. So, for example, if we notice a characteristic pattern of correlations among properties of objects, say that long-haired animals have slower metabolisms, and are more likely to be prey, we may infer that there is a common cause of all these properties, even if initially we aren't sure what that cause actually is. Similarly, when several members of an existing kind, like cats, turn out to have a particular new property, say they are immune to foot-and-mouth disease, we may infer that the common essence that caused this property in one cat will also be the cause of the property in the other cats, even if we cannot specify that causal relation in detail.

Another consequence of the causal view is that the effects of an object's "essence" will not be limited to properties of the objects themselves. The causal powers of the object may be responsible for the intrinsic properties of the object, but they will also be responsible, more widely, for causal effects on other objects or events. On our view, then, children and adults should pay special attention to the broader causal effects of objects when they categorize them. Notice that this view contrasts with the view that children or adults are simply registering patterns of correlation among the object properties themselves (see e.g., Jones & Smith, 1993). On our view, children and adults do pay attention to correlations among properties but their categorization is more than just a summary of those correlations. Instead, they use those correlations to infer causal powers. Those powers may have effects that extend well beyond the object itself.

In fact, inductions about causal effects on other objects play a particularly important role in the philosophical arguments in favor of natural kinds. For example, consider Putnam's classic "Twin Earth" arguments (Putnam, 1975). The stuff that runs in waters and streams on earth and on the far-away planet Twin Earth are perceptually indistinguishable, and so my twin on Twin Earth and I will have associated the same perceptual experiences with "water." Nevertheless, unbeknownst to us, Twin Earth water really has a very different chemical composition than water, it is made of xyz rather than H2O. Our intuition in this case, according to Putnam, is that the words refer to different kinds of things, in spite of the fact that they are associated with the same perceptual features. How does this intuition translate into an empirical prediction? The prediction is that some causal effects of xyz will turn out to be different from the causal effects of H2O. This is how the scientists will eventually discriminate between them. In fact, in the scientific case, these new predictions will often turn out to involve causal relations among objects rather than intrinsic properties. H2O and xyz may register differently on a spectrometer, for example, or react differently with other chemicals.

We argue, then, that in our kind understanding we assume that perceptual features are important only in so far as they are an indicator of underlying causal powers. Hence, objects that look similar but appear to have different causal powers will be considered to belong to different kinds and will receive different names. Conversely, the fact that perceptually similar objects receive different names will be a cue to the fact that they belong to different kinds and have different causal powers.

This interpretation of kind reasoning as a species of causal reasoning makes a great deal of functional and evolutionary sense. Often, in fact, usually, the causal structure of the world and the causal powers of objects are not immediately apparent. TV remotes make the set turn on, magnets make filings move towards them, water makes a plant grow, the sun makes you get sunburned, a snake makes you scared, but none of these causal relations are perceptually obvious. We can't simply predict how an object will affect other objects by seeing what it looks like. In fact, similar looking objects may have different causal effects on other objects, the stereo remote doesn't turn on the TV, ordinary metal doesn't attract filings, vodka kills plants, indoor lights don't burn you, garden hoses aren't scary. Somehow we must learn about these causal powers of objects, determine which objects do or do not have those powers, and name them accordingly.

Moreover, understanding hidden causal powers, and categorizing objects in terms of them, is extremely useful. It allows us to make important predictions about what will happen in the future, through the process of causal induction. It also allows us to intervene in the world effectively to make things happen ourselves. We can make TVs turn on or plants grow or scare our relatives. By putting objects in the same group or giving them the same name we are making a powerful prediction about their current and future causal powers.

Even more important, assuming that names refer to kinds allows us to quickly and efficiently learn about the causal powers of objects from others, and to communicate our own knowledge about causal powers to them. A child (or an adult) who hears that two objects have the same name knows much more than just that the objects look the same (which is obvious in any case). She also knows that the objects' properties have a common cause and that this will lead to common effects in the future.

The theory theory, causal maps and Bayes-Nets

The idea that kinds involve common causal powers is one part of our broader conception of the nature of cognition and cognitive development - and, in particular, our interpretation of "the theory theory." The theory theory is the idea that much of our adult knowledge, particularly our knowledge of the physical, biological and psychological world, consists of "intuitive" or "naïve" or "folk" theories (Murphy & Medin, 1985; Rips, 1989). Similarly, cognitive developmentalists argue that children formulate and revise a succession of such intuitive theories (Carey, 1985; Gopnik, 1988; Keil, 1989; Wellman, 1990; Gopnik & Meltzoff, 1997; Wellman & Gelman, 1997). This idea rests on an analogy between everyday knowledge and scientific theories. Advocates of the theory theory have drawn up lists of features that are shared by these two kinds of knowledge (see e.g., Gopnik & Wellman, 1994). The assumption behind this work has been that there are common cognitive structures and processes, common representations and rules, that underlie both everyday knowledge and scientific knowledge.

Formulating the analogy between the history of science and development has been an important first step, but it is time to try to describe in some detail the representations and rules that could underpin both these types of knowledge. In order to specify both the idea of theories and of kinds more precisely, it is necessary to flesh out the nature of those cognitive structures and processes in more detail. Ideally, such an account should include ideas about the computational character of these representations and rules.

We have recently begun to outline a more developed cognitive and computational account of the theory theory. In particular, we argue that many everyday theories and everyday theory changes involve a type of representation we call a "causal map" (Gopnik, 2000; Gopnik & Glymour, in press). A causal map is an abstract, coherent, learned representation of the causal relationships among kinds of objects and events in the world. These maps are analogous to the spatial cognitive maps used by many animals (Tolman, 1932; O'Keefe & Nadel, 1978).

These representations are different from some more familiar kinds of causal knowledge. For example, there is some evidence that certain kinds of specific causal inferences, such as inferences about the interactions of moving objects, are in place very early and may even be hardwired. For example, there is evidence that adults and children assume that when one object collides with another object it will cause the second object to move. Michotte (1962) showed that adults who see such abstract movement patterns of objects in an animated film automatically interpret them as causal interactions. He interpreted this as evidence for the "perception" of causality. Following Michotte, developmentalists have shown that infants who witness such movement patterns show distinctive looking-time patterns. In particular, infants who witness one object collide with another seem to expect that the second object will move (Leslie, 1982; Oakes & Cohen, 1995).

In addition, familiar types of learning such as classical and operant conditioning may also be seen as involving implicit causal inferences. Many modern investigators of animal cognition suggest that these types of learning are effective because they capture the causal character of events (Shanks & Dickinson, 1987; Gallistel, 1990).

Causal maps, however, go beyond either of these types of causal knowledge. Unlike the perceptual Michottean principles, these maps do not simply involve a few basic notions of folk physics but may involve a wide variety of novel causal relations. And unlike the Michottean principles, causal maps may be learned. Unlike classical conditioning, these relationships do not just involve a few ecologically significant events such as food or pain. Unlike operant conditioning, they do not just involve the organism's own actions, but rather involve judgments of the causal relations among objects and events in the world that are independent of our actions upon them.

"Everyday" or "folk" theories seem to have much of the character of causal maps. Such everyday theories represent causal relations among a wide range of objects and events in the world independently of the relation of the observer to those actions. They postulate coherent relations among such objects and events which support a wide range of predictions, interpretations and interventions. Moreover, theories, like causal maps, are learned through our experience of and interaction with the world around us.

The idea of causal maps also seems to capture the scope of "theory theories" very well. The theory theory has been very successfully applied to our everyday knowledge of the physical, biological and psychological worlds. However, the theory theory does not seem to be as naturally applicable to other types of knowledge, for example, purely spatial knowledge, syntactic or phonological knowledge, musical knowledge or mathematical knowledge. Nor does it apply to the much more loosely organized knowledge involved in empirical generalizations, scripts or associations (Gopnik & Meltzoff, 1997). But these types of knowledge also do not appear to involve causal claims in the same way as folk physics, biology and psychology. Conversely some kinds of knowledge that do involve causal information, like the kinds of knowledge involved in operant or classical conditioning, do not seem to have the abstract, coherent, non-egocentric character of causal maps, and we would not want to say that this sort of knowledge was theoretical.

On this view, kinds, in everyday understanding, would be the sort of entities that appear in causal maps. In fact, the standard view in philosophy of science is that kinds in scientific theories are defined by the causal relations they enter into, what is sometimes called their "causal role". Gold, from a scientific point of view, is not gold because it glitters but because it has a particular, regular set of causal interactions with other elements, regularities that are explained by atomic theory. If the analogy between science and everyday understanding is correct, everyday kinds also should be defined by their causal role.

Bayes Nets

We have proposed, then, that children and adults construct causal maps: non-egocentric, abstract, coherent representations of causal relations among objects and events. There has recently been a great deal of computational work investigating similar representations. The representations commonly called "Bayes nets" can model complex causal structures and generate appropriate predictions and interventions. A wide range of normatively accurate causal inferences can be made, and, in many circumstances, they can be made in a computationally tractable way. The Bayes net representation and inference algorithms allow one sometimes to uncover hidden unobserved causes, to disentangle complex interactions among causes, to make inferences about probabilistic causal relations and to generate counterfactuals see Spirtes, Glymour, & Scheines, 1993; Jordan; 1998; Pearl, 1988, 2000; Glymour, & Cooper, 1999; Spirtes, Glymour, & Scheines, 2000; Glymour, in press).

This work has largely taken place in computer science, statistics and philosophy of science. But these computational theories might also provide important suggestions about how human beings, and particularly young children, represent causal information. Causal maps might be a kind of Bayes net.

The details of the Bayes net formalism are complex, but the basic logic is quite simple and familiar. Causal relationships between events are systematically related to patterns of conditional probability among those events. This fact allows us to use causal knowledge to make predictions about probabilities, and it also allows us to use probabilities to make inferences about causal relations. Take the very simplest case, where x and y are the only events. If we know that x causes y and we observe that x occurs, then we can predict that y will be more likely to occur. Similarly, if we know that x causes y and we want y to occur, we can intervene to make x occur and assume that y is likely to follow. Finally, if we know that when x occurs y is likely to occur, that is, we know that x and y are correlated, we can work backwards and infer that there is a causal relation between them.

For example, suppose x is drinking wine, and y is insomnia. If I already know that drinking wine causes insomnia I can predict that if I drink wine I'll be more likely to stay awake. If I want to stay awake I should drink wine, and if I want to avoid insomnia I should avoid wine. On the other hand, if I simply observe that when I drink wine I am more likely to stay awake, I may work backwards to infer that drinking wine causes my insomnia..

The trouble, of course, is that this simple case is too simple. Introducing just one more type of event makes the story a lot more complex. For example, suppose we introduce z , going to parties. X, y and z could be related in a number of different ways. For example, z could be a common cause of x and y. Going to parties could make me more likely to drink wine, and it could also, independently, make me more likely to stay awake from sheer excitement. On the other hand, z could cause x which in turn could cause y, in a causal chain. Going to parties could make me more likely to drink wine which could make me more likely to stay awake. (see Fig. 1)

These two different causal structures will lead to very different predictions about the conditional probabilities of the events and they will lead to different interventions. If I know that the right causal structure is A , the common cause, for example, I would predict that I will stay awake after a party, even if I don't drink. If I know that it is B, the causal chain, I would predict that I will only stay awake after a party if I do drink. Similarly, my techniques for avoiding insomnia will be quite different in the two cases.

Conversely, different patterns of conditional probabilities among the three events will allow me to infer different causal structures. Simply seeing that x and y are correlated won't tell me which of these two causal structures is correct - hence, of course the mantra that correlation doesn't imply causation. However, I can make the right inference if I consider whether x is correlated with y conditional on z. For example, if I observe that wine and insomnia are only correlated when I also go to a party I may conclude that A is right. If I observe that wine and insomnia are correlated whether or not I go to a party, I may conclude that B is right.

Bayes nets generalize and formalize this kind of reasoning about the relations between conditional probabilities and causal structure. Bayes nets are directed graphs, like the ones in Fig. 1 or the more complex graph in Fig. 2. The nodes of the graph represent variables, whose values are properties of the system to which the net applies. "Color," for example, might be a variable with many possible values; "weight" might be a variable with two values, heavy and light, or with a continuum of values. When Bayes nets are given a causal interpretation, a directed edge from one variable to another, X toY, for example, says that an intervention that varies the value of X but otherwise does not alter the causal relations among the variables will change the value of Y. In short, changing X will cause Y to change.

In a causal Bayes-Net we assign a probability to each value of each variable, subject to a fundamental rule, the causal Markov assumption. The Markov assumption says that if the edges of the graphs represent causal relations, then there will only be some patterns of conditional probabilities of the variables, and not others. It is a generalization of the sort of reasoning we used in the wine/insomnia example. The Markov assumption constrains the probabilities that can be associated with a network. In particular, it says that the various possible values of any variable, X, are independent of the values of any set of variables in the network that does not contain an effect (a descendant of X), conditional on the values of the parents of X. So, for example, applied to the directed graph in Figure 2, the Markov assumption says that the value of X is independent of {R, Z} conditional on any values of variables in the set {S,Y}.

Bayes nets allow us to make causal predictions from observations. Information that a system has some property or properties often changes the probabilities of other features of the system. In the example above, for example, if I know that wine causes insomnia and I know that you have been drinking, that makes it more likely to be true that you will have insomnia. Conversely, if I know that you have insomnia that makes it more likely that you had been drinking earlier. Such changes are represented in Bayes nets by the conditional probability of values of a variable given values for another variable or variables. Bayes net representations simplify such calculations in many cases. In the network above, the probability of a value of X conditional on a value of R may be calculated from the values of p(R | S), p(S) and p(X | S) (See Pearl, 1988). This allows us to predict the value of X if we know the value of R. Bayes-nets also provide a natural way of assessing and representing counter-factual claims (Pearl, 2000).

When we intervene in the world we specifically predict the outcome of an action. The probabilities for various outcomes of an action that directly alters a feature are not necessarily the same as the probabilities of those outcomes conditional on that altered feature. Simply observing is not the same as intervening. In our example, for instance, if I observe that you are awake now that will make it more likely that you drank wine last night, but if I intervene to keep you awake now that won't make it more likely that you drank wine last night.. Suppose R in the graph above has two values, say red and pink. Because the value of S influences R, the conditional probabilities of values of S given that R = red will be different from the conditional probabilities of values of S given that R = pink. Because S influences X, the probabilities of values of X will also be different on the two values of R. Observing the value of R gives information about the value of X. But R has no influence on S or X, either direct or indirect, so if the causal relations are as depicted, acting from outside the causal relations represented in the diagram to change the value of R will do nothing to change the value of S or X. It is possible to compute over any Bayes network which variables will be indirectly altered by an action or intervention that directly changes the value of another variable. (See Spirtes, Glymour, & Scheines, 1993; Glymour & Cooper, 1999).

Bayes-nets thus have two of the features that are needed for applying causal maps: they permit prediction from observations, and they permit prediction of the effects of interventions. With an accurate causal map, that is the correct Bayes-net representation, we can accurately predict that y will happen when x happens, or that a particular change in x will lead to a particular change in y, even when the causal relations we are considering are quite complex. Similarly, we can accurately predict that if we intervene to change x then we will bring about a change in y.

This formalism thus provides a natural way of representing causal relations, and it allows for their use in prediction and intervention. The formalism also allows us to work backwards from patterns of conditional probability to causal structure – it allows learning. Given a particular set of conditional probabilities among variables the formalism can tell us which causal structures are compatible with those probabilities. Moreover, a variety of algorithms have been designed that do this in a computationally efficient way. (Glymour and Cooper, 1999 ; Jordan , 1998). In fact, some algorithms will correctly infer hidden unobserved causal variables based on the patterns of correlation of observed variables. In recent work in artificial intelligence, systems using Bayes-nets can infer accurate, if often incomplete, accounts of causal structure from suitable correlational data in fields ranging from epidemiology to mineralogy to genetics.

The Bayes-Net formalism, then, provides a way of representing and computing normatively accurate causal predictions and interventions and also provides algorithms for learning causal structure from data. The last thirty years of research in developmental psychology have taught us that even very young children know a great deal about the causal structure of the world, and can generate causal predictions and interventions in a wide range of domains. Some of this knowledge may be innate, but children also seem to learn an almost incredible amount about the causal structure of the world in a relatively short time. In fact, this has been one of the major contributions of work in the theory theory. We have proposed that children may use more heuristic implicit versions of the Bayes-Net computational tools to accomplish these feats (Gopnik & Glymour, in press; Gopnik, et al, in press).

How could we represent the idea of kinds in this sort of formalism? We may think of kinds as a certain subset of variables in a causal graphical model. This subset has two characteristics. First, a kind variable is the common cause of a set of other variables. Second, those other variables are all properties common to a group of objects. So to say that an object is a member of the kind "tiger" is to say that there is a variable called "tiger" which is a common cause of other values of other properties of the same object (e.g., having stripes, being a predator, growling), and is also a common cause of those properties of other objects of the same category/kind. To say that an object is a member of a kind is to say that its features, and similar features of other members of the kind, have a common cause in a relevant causal graph. Reasoning about kinds, on this view, is a particular type of causal reasoning.

Notice again, that this kind of representation if kinds would allow for two kinds of kind reasoning. First, we could use our existing causal knowledge to make kind inferences. We could consult our causal graph and see if there is a common cause for a set of object properties, and if so, what that cause is likely to be. But we could also draw conclusions about kind membership by simply looking at the pattern of correlations among object properties. If those patterns are consistent with a "common cause" structure, we might posit a common "essence" that caused the properties, even if we had no prior knowledge of that causal relation.

Naming, Sorting and Causal Induction in 2-to-4 year olds - The blicket detector

How do these theoretical ideas translate into an empirical research program? We predict that children will categorize objects together, and give them the same name, when they believe that the objects have common causal powers. Similarly, when children see others categorize objects together or give them the same name, they will assume that those objects have common causal powers. In fact, there is some evidence that adults and even older children categorize kinds in terms of causal powers in just this way (Ahn et al., 2000; Rehder, in press). Rheder has demonstrated empirically that they do so in a way that is predicted by the Bayes-Net formalism. But, of course, adults have extensive experience and often explicit training in causal inference. We predict that this will also be true of much younger children.

In a series of recent experiments, we "invented" a brand new causal power of objects, whether or not they made a particular machine work. We constructed a new machine, the "blicket detector:" a box that lights up and plays music when certain objects, but not others, are placed upon it. This apparatus presents children with a new, never encountered causal relation, of which they are given direct experience with real objects. Since we had complete control over this property we could also directly pit it against perceptual features of objects. We could arrange to have perceptually identical objects that did or did not display the causal power. We then explored whether children would use the causal power as a basis for naming the objects and whether children would use names to guide their inductions about this causal power. We also explored how these behaviors were related to perceptual similarities and differences among the objects.

In these experiments, children are presented with a 5" x 7" x 3" box, made of wood with a red lucite top. Two wires emerge from the detector's side. One is plugged into an electrical outlet. The other runs to a switchbox. If the switchbox is in the "on" position, the detector will light up and play music when an object is placed upon it. If the switchbox is in the "off" position, the detector does nothing when an object is placed upon it. During the experiment, this wire runs to a confederate who surreptitiously flips the switch on to allow an object to set the machine off, or flips it off to ensure that an object will not set the machine off. The wire and switch-box are hidden from the children's view and they have no suspicion of the role of the confederate, whom they never see. The apparatus is designed so that when the switch is on, the box "turns on" as soon as the object made contact with it and continues to light up and play music as long as the object continues to make contact with it. It "turns off" as soon as the object ceases to make contact with it. This provides a strong impression that something about the object itself caused the effect. The "objects" can vary according to the experiment and include wooden blocks as well as more "natural" objects like small rocks and hardware parts. We will report the results with blocks, but in fact we obtained identical results with both types of objects. (see Fig.3).

In one experiment (Gopnik & Sobel, 2000, Experiment 1), 2, 3- and 4-year-olds were shown four wooden blocks. The experimenter placed each block on the machine in succession, and then carefully returned it to its original location. Two blocks set the machine off and two did not. After this was demonstrated twice, children were told that one of the objects that had set the machine off was a "blicket." Then the experimenter asked the child to give him the other "blicket." Importantly, children were not told that the machine was a blicket detector and had no prior exposure to this novel causal property.

Children were given two types of sets of blocks, neutral sets and conflict sets. In the neutral sets, the blocks were either all identical, or all different. In the conflict sets, there were two pairs of perceptually identical objects, one member of each pair would set the machine off and the other would not. Here, the perceptual features of the object actually conflicted with the objects' causal powers. (see Fig. 3)

In the neutral tasks, even the 2-year-olds categorized the object on the basis of its causal power. In the conflict tasks, 2-year-olds chose as the other "blicket" the perceptually similar object more often than the causally similar object, but still chose the causal object more frequently than a distractor object. The 3- and 4-year-old children were equally likely to choose the causally or perceptually similar object as the "blicket" (see Fig. 4).

In a control condition (Gopnik & Sobel, 2000, Experiment 2), the same machine and objects were presented to other children of the same ages with the same procedure. However, in this condition the object did not appear to be causally related to the machine. Instead of placing each object on the machine, the experimenter would hold each object over the machine. For two of the objects, he would simultaneously press the top of the detector with his hand, which activated it. For the other two, he simply held his hand near the top of the detector, but did not press it and nothing happened. Children were then told that one of the blocks which had been associated with the machine's activity was a "blicket" and were asked to show the experimenter the other "blicket." In contrast to the first experiment, children of all ages chose at chance in the neutral tasks and used the perceptual properties of the object as a basis for categorization in the conflict tasks. Children would not categorize an object as a "blicket" based on a mere association between that object and the machine's activation by the experimenter (see Fig. 5). They would only categorize when there appeared to be a causal relation between the block and the machine.

In a second condition, we showed that children would not only use causal powers as a guide to names, but would also use names as a guide to causal powers. In this experiment, we told children that two of the objects were called "blickets" and two were not. The experimenter then picked up one of the "blickets" and showed that it activated the machine. The children were asked to choose "another one that will make the machine go." Thus this experiment was analogous to Gelman et als.. induction studies, but with real objects displaying real novel causal powers. Again 2-, 3- and 4-year-olds predicted that the objects with the same name would have the same causal powers in the neutral condition. The 3- and 4-year-olds clearly did this even in the conflict condition when this prediction conflicted with a perceptual categorization (e.g., when one "blicket" was a green square and the other was a red cylinder, while the two non-"blickets" were also a green square and a red cylinder). This effect was less strong for the 2-year-olds (see Fig. 6).

It is important to say that children, particularly 2-year-olds, did sometimes use perceptual features as a basis for categorization or induction in the conflict conditions. However, it turned out that these children also tended to misremember the causal information or the object's name. Many of the 3- and 4-year-olds, and most of the 2-year-olds, who said that the perceptually identical object was the "blicket," also said, incorrectly, that that object had set off the detector. Similarly, when they predicted that the perceptually identical object would activate the detector, they also tended to say, incorrectly, that that object had been called a "blicket." That is, if they said that the two green squares were both "blickets" they were also likely to say, incorrectly, that both green squares had set off the machine. Conversely, if they said that the two green squares would both activate the machine, they were also likely to say, incorrectly, that they had both been labeled as "blickets."

This suggests that names and causal powers may have been linked even for children who made a perceptual response. These memory errors may suggest that children, and particularly the younger children, assume that perceptual, causal and linguistic properties are correlated. In fact, of course, this is usually true, and the inference from common features to common causal powers is correct. Most of the time when objects have the identical shape, color, and size they also have the same causal powers. For these children, this fact guides not only their categorizations and inductions, but also their memory. They seem to assume that the identical objects <u>must</u> have the same causal powers, and simply ignore the data to the contrary. In contrast, if children were simply using a purely perceptual strategy to categorize, name, and make inductions about objects, these errors are puzzling.

In a second study, we set out to test this hypothesis about memory errors more systematically (Nazzi & Gopnik, 2000). In earlier studies, investigators in the shape vs. kind debate had shown that either perceptual or non-perceptual cues may be used in categorization, but no one had systematically pitted these cues against one another. The Gopnik and Sobel study suggested that younger children might rely more heavily on perceptual cues than older children but there were no significant age effects, only a trend. Moreover, it was possible that the memory errors occurred simply because children were confused about the task. In this study, we followed exactly the same procedure as in Gopnik and Sobel but added a memory control task at the start of each session: children were simply asked to report which object had set off the machine, without also categorizing the objects. All the children did well on this task. We also used a smaller range of children at each age, comparing 3½ and 4½-year-old children born within a 2-3 months range. This allowed us to reveal an age effect, with younger children making more perceptual responses than older children.

The memory errors showed that almost all of the age effect was due to the fact that younger children were more likely to assume a correlation between perception and causation than older children. Children in each age group said that the perceptual object was a "blicket" but correctly remembered that it had not set off the machine about 20% of the time. However, the younger children falsely reported a correlation between the perceptual and causal features of the objects (e.g., they said that the perceptually similar object both was a "blicket" and had set off the machine before) 46% of the time. The older children only made this error 12% of the time.

This suggests that the developmental effect is not simply due to a shift from perceptual to causal responding. Rather the younger children appeared to treat perceptual cues as a guide to causal ones, to the point of misremembering cases were those cues are not correlated. This may also help to explain the debate in the literature. In real life, perceptual cues, and particularly shape, are highly correlated with causal powers. Other things equal, objects with the same shape are also likely to have similar causal powers. Children may, in fact, use shape to categorize and name objects. However, they may do so almost entirely because for them shape is a good predictor of causal powers. When the data shows them that shape and causal powers conflict they either reinterpret the data (at three) or prefer causal powers (at four).

Sorting, naming and causal powers

So far we have talked about two indices of categorization, naming and causal induction, and we have shown that these two phenomena are closely related in children as young as 30 months old. When children know that objects have the same causal powers they predict that they will have the same name and vice-versa. Another, non-linguistic, index of categorization is manual sorting, children will put similar objects in the same location. How are these three indices of categorization -- sorting, naming and causal induction, related to one another?

We designed a simplified "blicket detector" paradigm to ask this question for 30-month-olds (Nazzi, 2001; Nazzi & Gopnik, in preparation). In this paradigm children saw the experimenter place three perceptually different objects on the detector, twice in a row. Two of the objects consistently activated the detector and one did not. The experimenter then picked up one of the active objects and presented the child with the other two objects. In the sorting condition, he visibly placed one of the active objects in one hand and asked "Can you give me the one that goes with this one?" Thus, in this condition, children had to put the causally similar objects in the same location, rather than to give them the same name. In the action condition, he asked the child to actually make the machine work ("Can you make it work?"), testing whether children really had inferred the causal power of the objects correctly. We also included a control condition, just like the control in the Gopnik and Sobel (2000) study we described above. In this condition, the experimenter held the objects up over the detector and then either pressed the top of the detector to activate it, or held his hand near the machine, but did not activate it.

Thirty-month-olds consistently used the causal powers of an object to sort the objects, just as they did to name the objects. 78% of the time they placed together objects that had made the detector work and 79% of the time they correctly predicted that the object would make it work again. This suggests that there is not only a link between causal induction and naming, but between causal induction and categorization more generally. Moreover, as in Gopnik and Sobel

(2000), 30-month-old children did not physically sort objects together or use them to activate the machine when the objects were simply associated with the machine in a non-causal way. They were at chance in both tasks. This again suggests that children are reacting to causal powers and not mere associations.

These experiments, then, demonstrate that children's understanding of the structure of kinds develops between 30 months and 4½ years of age, by which time they have reached something similar to the adult conception. In particular, we have converging evidence that children this young will give perceptually different objects the same name, and sort them together, when they have the same causal powers. Similarly, they will use naming as a guide to causal induction even without perceptual cues. Even in cases where perceptual categorization conflicts with causal categorization, the younger children will sometimes, and the older children will usually, rely on causal cues. And even when, in conditions of conflict, perceptual cues are used, those cues seem to be correlated with causal powers. Moreover, and importantly, children will not just categorize on the basis of any type of similarity. When objects are merely associated with effects but do not actually cause them, children will not use those associations as a basis for naming or sorting the objects.

In fact, the crucial developmental change in this period appears to be a loosening of initially tight correlations between perception and causation. The younger children seem more inclined to believe that perceptual, linguistic and causal similarities will be correlated, while the older children seem more willing to imagine that perceptual cues could conflict with linguistic or causal cues.

Naming and sorting in 18 month olds

What about even younger children? In a further experiment with "the blicket detector," we did not find similar effects with 24-month-olds (Nazzi, 2001; Nazzi & Gopnik, in preparation).

This seems to go against the idea that younger children have a similar understanding of kinds and causal powers. However, this conclusion might be too strong. First, it cannot be totally excluded at this point that the failure of the 24-month-olds is due to the attention and memory demands of this task. Second, in other studies, we have explored two abilities that do emerge in younger children and that may be indicators of, or at least prerequisites for, "kind" understanding. These are the ability to sort all the objects in a group into multiple categories and the ability to sort perceptually dissimilar objects together when they receive the same name.

A number of studies dating back to the seventies demonstrate some important changes in children's sorting behavior at about 18 months. From about a year of age, children will categorize objects by placing similar objects in the same location.(Ricciutti, 1965; Nelson, 1973; Starkey, 1981; Sugarman, 1983; Gopnik & Meltzoff, 1987, 1992;). In these studies, children are given a mixed group of objects (e.g., four identical yellow rectangles and four identical clear pill-boxes) and are allowed to spontaneously manipulate those objects. Children as young as nine months of age will place the similar objects in one group in the same location. However, while this behavior might be indicative of categorization, it might also just reflect a preference for one type of object over another. It is only at about 18 months that children will systematically and exhaustively sort all the objects in a group into separate locations. We demonstrated that children will also do this when the objects in each group are not identical but belong to the same "basic-level" category, using for example a set of four different rings and four different pencils. We also suggested that this ability might be related to an understanding of "kinds" (Gopnik & Meltzoff, 1992)..

We also demonstrated that there was a close empirical connection between this sort of exhaustive sorting and naming. In longitudinal studies, children began to systematically and exhaustively sort objects into multiple categories (Gopnik & Meltzoff, 1987) a week or so before they had a "naming spurt" -- a sudden sharp increase in their naming vocabulary. In cross-sectional studies, similarly, the ability to exhaustively sort objects into multiple categories was highly correlated with the number of names children used, and with their mothers' report of a naming spurt (Gopnik & Meltzoff, 1987, 1992). Moreover, this relationship was quite specific, it remained strong even when age was held constant, and there was no parallel relation between the naming spurt and other cognitive developments that take place at about the same time such as means-ends and object permanence achievements (Gopnik & Meltzoff, 1992). Mervis and Bertrand (1994, 1995, 1997) replicated this finding both with typically developing children and children with Down's syndrome. Interesting, they did not find this relation in children with Williams' syndrome who appear to approach language development in a way that is very different from these other groups. Mervis and Bertrand (1994) also showed that exhaustive sorting was related to the achievement of "fast mapping:" Children who produced exhaustive sorting were also more likely to quickly learn to attribute new words to unnamed objects.

The fact that the naming spurt and the new categorization abilities occurred so closely together suggested an intriguing possibility - perhaps the emergence of the new linguistic ability to learn names itself facilitates a cognitive change in children's sorting abilities. In fact, this relation between naming and sorting parallels two other independent relations between language and cognition in this period. In even earlier work, we demonstrated parallel relations between the acquisition of words for disappearance, such as "allgone," and object-permanence abilities, and of words relating to action, such as "uh-oh," and means-ends abilities (Gopnik, 1982, 1984, Gopnik & Meltzoff, 1984). We have argued that in all these cases the linguistic and cognitive abilities may emerge in tandem and be mutually facilitating – the cognitive developments spark the linguistic acquisitions and the linguistic developments also strengthen the cognitive changes (Gopnik & Meltzoff, 1997).

Particularly dramatic evidence for this sort of "neo-Whorfian" link between naming and sorting comes from cross-linguistic studies. Korean relies much more heavily on verb morphology than English and often omits nouns. In a series of studies, we found that Korean-speaking mothers used significantly fewer nouns in their speech to children than English-speaking mothers (Gopnik & Choi, 1990, 1995; Choi & Gopnik, 1995; Gopnik, Choi, and Baumberger, 1996). In turn, Korean-speaking children used significantly fewer nouns and more verbs in their own early language than English-speakers, a result replicated by Tardif et al. (1999) for Mandarin-speakers. In fact, the Korean-speakers, and also Mandarin-speakers, often used verbs at the same time or before they used nouns, a definitive refutation of the widespread idea that nouns precede verbs in development.

More significantly for the current argument, however, the cognitive development of the two groups of children also differed (Gopnik & Choi, 1990; Gopnik et al., 1996). The Korean-speakers began to produce exhaustive sorting significantly later than the English-speakers. Just as importantly, however, this effect was not due to some across-the-board cognitive delay. The Korean-speaking children were actually advanced in means-ends understanding compared to the English-speakers. Means-ends abilities, which involve an understanding of skilled goal-directed action, are conceptually more closely related to the verbs that play a prominent role in early Korean vocabularies. This cross-linguistic result suggested a quite specific link between the experience of hearing many new names and the tendency to sort objects into groups. The most plausible explanation for this finding was that the linguistic differences in the input to the children actually caused the cognitive differences. Hearing new names might lead children to categorize and sort objects in a new way.

In a more recent experiment (Nazzi & Gopnik, 2001), we tested this idea further by exploring whether hearing that objects were given the same name would actually lead children to sort those objects together. Could we demonstrate an effect of naming on sorting more directly? Sixteen- and 20-month-olds were given triads of objects that were completely different perceptually. In one condition, two of the objects were given the same name by the experimenter and a third object was given a different name. Then, just as in our earlier sorting experiment, the experimenter put one of the two similarly-named objects in his hand and asked "Can you give me the one that goes with this one?. Twenty-month-olds, but not 16-month-olds, physically sorted the objects with the same name together.

Moreover, this ability was related to vocabulary size. Not only did the 20-month-olds have significantly larger vocabularies than the younger children, but among the 20-month-olds vocabulary size was highly correlated with performance on the sorting task. This suggests that the capacity to sort objects into entirely new non-obvious categories based on names may itself be related to the naming spurt.

In a control condition, children received a visual categorization task, in which two of the objects were perceptually similar, and no names were used. As one might expect 16-month-olds and 20-month-olds performed equally well on this task and there was no correlation to vocabulary size.

In summary, then, both the capacity to sort objects exhaustively into multiple categories and to sort perceptually dissimilar objects together on the basis of a common name seem to emerge at around 18 months. Both these abilities seem to be related to children's own productive naming vocabulary. Moreover, there is evidence, from the cross-linguistic studies, that the experience of hearing names may actually contribute to these sorting abilities.

Both these abilities also seem to be related to kind understanding. In multiple-category sorting children seem to go beyond simple perceptual generalizations. Instead, these children seem to assume that all objects belong in some category or another. The capacity to sort completely dissimilar objects into a common category is even more obviously related to kind understanding. Indeed, the very nature of kind understanding involves this ability to go beyond perceptual similarities and dissimilarities and to categorize objects together in a more abstract way.

We do not know how these early sorting abilities are related to an understanding of causal powers. In Gopnik & Sobel (2000) and Nazzi (2001), we did not find the first signs of causal categorization until 30 months. It may be that the earlier types of understanding are prerequisites for a causal powers view of kinds, or it may be that such a view emerges in tandem with these abilities, and we have not yet designed a way to test it. Some support for the first alternative comes from the Gopnik and Sobel (2000) finding that it was easier to predict causal powers from the names given to objects, than to give the same name to objects having similar causal powers. It is possible that children's experiences with language, particularly the experience of hearing the same name applied to perceptually dissimilar objects, itself leads them to look beyond perceptual similarities, and that this, in turn, leads them to the idea of categorizing objects in terms of causal powers.

Kind understanding before 18 months.

Thus, we see development between 30 months and 4½ years leading to something very much like full-blown kind/causal power understanding, and we see some of the first signs of an understanding of kinds emerging at 18 months in tandem with the emergence of a naming spurt. What about even younger, preverbal, infants? Here, the picture seems to us much murkier. It is clear that, from birth, infants can individuate and identify objects. Like other organisms, infants can also discriminate among objects perceptually and can make perceptual generalizations.

It is also clear that, at least from very early in development, and quite possibly from birth, infants can make certain kinds of causal predictions and interventions, and can learn about new causal relations. This capacity seems clear from the extensive literature on young infants understanding of principles of folk physics (see e.g., Bower, 1974; Baillargeon, 1995; Leslie, 1982; Oakes & Cohen, 1995; Spelke et al., 1992), and also from the earlier literature on infants understanding of contingency (Watson, 1987) and the emerging literature on infants' understanding of the actions of others (Meltzoff, 1995; Woodward, 1998; Gergeley, 1995). It seems that very young infants know that certain objects will have consistent causal effects on others in regular ways, and they also know that their own actions and the actions of others will have certain regular causal effects (see also Younger and Rakison, this volume).

We do not know, however, when children begin to combine their perceptual categorizations of objects and their causal knowledge. When do children first conceive of categories of spatially individuated, distinct, objects that nevertheless have common underlying causal powers, and that will have similar causal powers in the future? Many of the early types of causal understanding do not seem to have this character. Rather than seeing object categories as the sorts of things that define causal powers, infants initially appear to focus on movement paths and trajectories, for objects, or patterns of intentional action, for people. Certainly, to our minds there is no clear evidence in early infancy that children go beyond perceptual similarities and differences in their object categorizations, though those similarities and differences may sometimes be quite general or abstract (as in the types of categorization reported by Mandler et al., 1995). On the other hand, this question has yet to be tested in an appropriate and systematic way.

There is some evidence that towards the end of the first year children do begin to link their perceptual categorizations and their causal inductions. This evidence comes from a study of children's exploratory behavior (Baldwin et al., 1993). In this study, 9-to-26-month-old infants saw an unexpected novel causal property of an object (e.g., a can that made a cow noise when it was turned over). Then they were presented with perceptually similar objects that did not exhibit this property. Children explored those objects in a way that suggested that they were surprised by this fact. They seemed to expect the original causal property to generalize to the new objects.

This kind of linkage between perceptual and causal similarity might be the first emergence of the sort of conception of kinds that we see quite clearly in 30-month-olds. Recall that these older children seemed to think that perceptual similarities and causal powers are tightly correlated. However, this linkage must, paradoxically, be loosened in order to develop a full conception of kinds.

Learning

We might propose then a three-stage understanding of the link between perceptual and causal categorization, producing a kind of u-shaped curve. The youngest infants might simply see no relation at all between perceptual object categorization and causal prediction. The older infants and young toddlers might begin to link these cues. They begin to believe that perceptual similarity will predict causal similarity. Only the older 3- and 4-year-olds, however, begin to believe that while this linkage usually holds, it can be broken, and this is due to deeper facts about the essences of objects.

An interesting possibility is that language itself is a major factor in these changes in children's developing understanding of kinds and causal powers. There is evidence that towards the end of the first year, linguistic cues may focus infants attention on perceptual similarities among objects (see e.g., Waxman & Markow, 1995; Woodward, Markman, & Fitzsimmons, 1994) . In these early cases, however, the language cues seem simply to be facilitating types of perceptual generalization that children can make independently. Nevertheless, this sort of linguistic effect might be the first step in a process of leading children to place objects together in a more abstract and profound way. We see this in our 20-month-old's ability to use names to categorize objects even in the absence of perceptual similarities at all (Nazzi & Gopnik, 2001). Similarly, in unpublished work, Baldwin has found that linguistic cues facilitate the exploratory causal inductions of children, at least, at 20 months of age (Baldwin, personal communication). Twenty-month-old children who heard that two partially similar objects were given the same name were more likely to predict that they would have the same novel non-obvious causal property.

In fact, children might realize that the link between perception and causal inference can be loosened precisely because of the fact that there is also a link between the names of objects and their causal powers. Learning names sets up a kind of three term relationship between perceptual similarities, causal powers and names. Names give children a way to state a causal power similarity that is decoupled from a perceptual similarity. We see the first step in this process in our 20-month-old's ability to use names to establish categories in a way that goes beyond perceptual similarity. By three or four, these children are also able to relate those name-based categories to causal powers in a way that completely bypasses perceptual similarity.

Similarly, the insight of the naming spurt, namely the realization that all things have names, may be related to the insight that all objects belong in categories, an insight that seems to underlie exhaustive sorting. Empirically, these two developments are closely related. Moreover, in the cross-linguistic work we have demonstrated a direct effect of language input on exhaustive sorting.

If we are correct that there are these developmental changes in children's understanding of kinds and causal powers, where do those changes come from? Is a causal powers view of categorization innate? If not, what learning mechanisms propel children to develop a conception of kinds and causal powers?

This last question, the question of mechanisms of learning, seems to us to be the crucial question we should be addressing in cognitive development in general. Fortunately, in the case of causal categorization, reasoning, inference, and induction, there are parallel developments in philosophy and computer science that may be particularly helpful. The formalism of graphical causal models, as we saw earlier, provides powerful, formal, techniques for making accurate causal inferences. In particular, within the Bayes-Net framework, it is possible to construct algorithms that infer causal structure from patterns of correlation among events. These algorithms are provably accurate and, at least in many cases, computationally efficient. The formalism provides powerful tools for reliably inferring causal structures from patterns of evidence.

We are just beginning to explore whether very young children actually use implicit versions of these computational learning procedures. In a recent set of experiments, we have discovered that children can determine the correct causal categorization of objects, that is, can determine which objects are "blickets" by employing some of the same basic axioms as these computational systems (Gopnik et al., in press). Two, three and four year old children could accurately infer the causal powers of an object by considering the patterns of variation and covariation between that object, other objects, and the "blicket" detector's behavior.

This sort of learning might also be responsible for some of the broader developmental changes in children's categorizations that we described. One way of thinking of the learning process is that children are driven to look for ever deeper and more powerful causal analyses of the world around them. There will be significant and informative patterns of correlation among the perceptual features of objects, the names they are given, and their causal effects. Children may be detecting and using these patterns of correlation to make causal predictions and inductions. In particular, children may reason backwards from the patterns of correlation among perceptual and linguistic features of objects and their effects on other objects to infer underlying causal powers of those objects.

Note that the claim is not that children are simply covariation detectors or that they match patterns of correlation in the input in their own behavior (as proposed by, Jones & Smith. 1993, Smith et al. 1996). Rather, we would argue that children use covariation information in combination with other kinds of knowledge, in order to uncover underlying causal structure.

Conclusion

We have argued that "kinds" can be usefully understood as groups of objects with common current and future causal powers. This conception stems from our general conception of intuitive theories as "causal maps": abstract representations of the causal structure of the world. In a series of empirical studies, we have shown that children do categorize and name objects based on their causal powers, from at least 30 months of age. Even younger children, around 18 months old, already show the ability to sort objects into multiple, exhaustive categories and to sort entirely perceptually dissimilar objects together when they are given a common name. These abilities do not appear to be present in younger infants. They also seem to be highly correlated with the acquisition of names and particularly with a naming spurt.

There also seem to be important changes in children's kind understanding. We suggest that the youngest infants may not relate perceptual similarities among objects and causal powers at all. At a later stage, from about a year until about three or four, children do seem to predict strong correlations between perceptual similarities and causal powers. However, with the acquisition of language, children may have the further realization that names are correlated with causal powers in a way that may bypass perceptual similarities. Eventually, this may be what leads them to a deeper conception of underlying essences.

Causal inference, induction and learning seem to us particularly promising avenues of investigation for several reasons. Empirically, we have discovered that causal understanding plays an important and deep role in a great deal of children's early knowledge, including their knowledge of categories. From a functional and evolutionary perspective, causal understanding is profoundly adaptive. And methodologically, the problem of causal inference is an area where powerful formal and computational tools and techniques are available. Such tools may help us make real progress towards a developmental science that is both genuinely developmental and genuinely scientific.

References

Ahn, W.-K., Gelman, S.A., Amsterlaw, J.A., Hohenstein, J., & Kalish, C.W. (2000). Causal status effect in children's categorization. Cognition, 76, B35-44.

Baillargeon, R.; Kotovsky, L.; Needham, A. (1995). The acquisition of physical knowledge in infancy. In Causal cognition: A multidisciplinary debate. Dan Sperber, David Premack, Eds;. Clarendon Press/Oxford University Press, New York, NY, p. 79-116.

Baldwin, D.A., Markman, E.M., & Melartin, R.L. (1993). Infants' ability to draw inferences about nonobvious properties : evidence from exploratory play. Child Development, 64, 711-728.

Bower, T. G. (1974). Development in infancy. San Francisco, Calif.:m W.H. Freeman.

Carey, S. (1985). Conceptual change in childhood. Cambridge, MA: MIT Press.

Choi, S. & Gopnik, A. (1995). Early acquisition of verbs in Korean: A cross-linguistic study. Journal of Child Language, 22, 497-530

Gallistel, C. R. (1990). The organization of learning. Cambridge, MA: MIT Press.

Gelman, S. A. & Bloom, P. (2000). Children are sensitive to how an object was created when deciding what to name it. Cognition, Aug, v76 (n2):91-103.

Gelman, S. A. & Coley, J. (1990). The importance of knowing a dodo is a bird: Categories and inferences in 2-year-old children. Developmental Psychology, 26, 796-804.

Gelman, S.A. & Markman, E. (1986). Categories and induction in young chidren. Cognition, 23, 213-244.

Gelman, S. A., & Wellman, H. M. (1991). Insides and essence: Early understandings of the non-obvious. Cognition, 38, 213-244.

Gergely, Gyoergy; Nadasdy, Zoltan; Csibra, Gergely; Biro, Szilvia. (1995). Taking the intentional stance at 12 months of age. Cognition, 56 (2):165-193.

Glymour, C. (in press). The mind's arrows: Bayes nets and graphical causal models in psychology. Cambridge, MA: MIT Press.

Glymour, C, & Cooper, G. (1999). Computation, causation, and discovery. Menlo Park, CA: AAAI/MIT Press.

Gopnik, A. (1982). Words and plans: early language and the development of intelligent action. Journal of Child Language, 9, 303-318.

Gopnik, A. (1984). The acquisition of gone and the development of the object concept. Journal of Child Language, 11, 273-292.

Gopnik, A. (1988). Conceptual and semantic development as theory change: the case of object permanence. Mind and Language, 3, 197-216.

Gopnik, A. (2000). Explanation as orgasm and the drive for causal understanding: The evolution, function and phenomenology of the theory-formation system. In F. Keil & R. Wilson (Eds.) Cognition and explanation. Cambridge, Mass: MIT Press.

Gopnik, A. & Choi, S. (1990). Do linguistic differences lead to cognitive differences? A cross-linguistic study of semantic and cognitive development. First Language, 10, 199-215.

Gopnik, A. & Choi, S. (1995). Names, relational words, and cognitive development in English and Korean speakers: nouns not always learned before verbs. In M. Tomasello and W.E. Merriman (Eds.), Beyond names for things: young children's acquisition of verbs (pp. 63-80). Hillsdale, NJ: Erlbaum.

Gopnik, A., Choi, S., & Baumberger, T. (1996). Cross-linguistic differences in early semantic and cognitive development. Cognitive Development, 11, 197-227.

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

Gopnik, A. & Meltzoff, A.N. (1984). Semantic and cognitive development in 15- to 21-month-old children. Journal of Child Language, 11, 495-513.

Gopnik, A. & Meltzoff, A.N. (1987). The development of categorization in the second year and its relation to other cognitive and linguistic developments. Child Development, 58, 1523-1531.

Gopnik, A. & Meltzoff, A.N. (1992). Categorization and naming: Basic-level sorting in eighteen-month-olds and its relationship to language. Child Development, 63, 1091-1103.

Gopnik, A., & Meltzoff, A. (1997). Words, thoughts and theories. Cambridge, MA: MIT Press.

Gopnik, A., & Sobel, D. (2000). Detecting bickets: How young children use information about novel causal powers in categorization and induction. Child Development, 71, 1205-1222.

Gopnik, A. & Wellman, H.M. (1994). The theory theory. In L.A. Hirschfeld and S.A. Gelman (Eds.), Mapping the mind: domain specificity in cognition and culture (pp. 257-293). New York: Cambridge University Press.

Gopnik, A., Sobel, D. M., Schulz, L. & Glymour, C. (in press). Causal learning mechanisms in very young children: Two, three, and four-year-olds infer causal relations from patterns of dependent and independent probability. Developmental Psychology.

Imai, M.; Gentner, D.; Uchida, N. (1994) Children's theories of word meaning: The role of shape similarity in early acquisition. Cognitive Development, 1994 Jan-Mar, v9 (n1):45-75.

Jones, S.S., & Smith, L.B. (1993). The place of perception in children's concepts. Cognitive Development, 8, 113-139.

Jordan, M. (Ed.) (1998). Learning in graphical models. Cambridge, MA: MIT Press.

Keil, F.C. (1989). Concepts, kinds, and cognitive development. Cambridge, MA: MIT Press.

Kemler-Nelson, D.G. (1995). Principle-based inferences in young children's categorization: revisiting the impact of function on the naming of artifacts. Cognitive Development, 10, 347-380.

Kemler Nelson, D.G., Frankenfield, A., Morris, C. & Blair, E. (2000a). Young children's use of functional information to categorize artifacts: Three factors that matter. Cognition, 77, 133-168.

Kemler Nelson, D.G., Russell, R., Duke, N., & Jones, K. (2000b). When young children name artifacts by their functions. Child Development, 71, 1271-1288.

Kripke, S. (1972). Naming and necessity. Cambridge, MA: Harvard University Press.

Landau, B., Smith, L.B., & Jones, S.S. (1988). The importance of shape in early lexical learning. Cognitive Development, 3, 299-321.

Leslie, A.M. (1982). The perception of causality in infants. Perception, 11, 173-186.

Mandler, J. M.; McDonough, L. (1993). Concept formation in infancy. Cognitive Development, , 8 3):291-318.

Meltzoff, A.N. (1995). Understanding the intentions of others: re-enactment of intended acts by 18-month-old children. Developmental Psychology, 31, 838-850.

Mervis, C.B. & Bertrand, J. (1994). Acquisition of the novel name-nameless category (N3C) principle. Child Development, 65, 1646-1662.

Mervis, C.B. & Bertrand, J. (1995). Acquisition of the novel name-nameless category (N3C) principle by young children who have Down syndrome. American Journal on Mental Retardation, 100, 231-243.

Mervis, C.B. & Bertrand, J. (1997). Developmental relations between cognition and language : evidence from Williams syndrome. In L.B. Adamson & M.A. Romski (Eds.), Research on communication and language disorders : contributions to theories of language development. New York, NY : Brookes.

Michotte, A. (1962). Causalité, permanence, et réalité phénoménales. Louvain: Publications Universitaires.

Murphy, G.L., & Medin, D.L. (1985). The role of theories in conceptual coherence. Psychological Review, 92, 289-316.

Nazzi, T. (2001). When do children start using causal properties to categorize objects? Poster presented at the Biennial meeting of the Society for Research on Child Development, Minneapolis, MN, April 19-22, 2001.

Nazzi, T. & Gopnik, A. (2000). A shift in children's use of perceptual and causal cues to categorization. Developmental Science, 4, 389-396.

Nazzi, T. & Gopnik, A. (2001). Linguistic and cognitive abilities in infancy: When does language become a tool for categorization? Cognition, 80, B11-B20.

Nazzi, T. & Gopnik, A. (in preparation). The emergence of the use of causal cues to object categorization.

Nelson, K. (1973). Structure and strategy in learning to talk. Monographs of the Society for Research in Child Development, vol. 38, no. 1-2 (serial no. 149). Chicago, MI: Society for Research in Child Development.

Oakes, L.M., & Cohen, L.B. (1995). Infant causal perception. In C. Rovee-Collier and L.P. Lipsitt (Eds.), Advances in infancy research (vol. 9, pp. 1-54). Norwood, NJ: Ablex.

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

Pearl, J. (2000). Causality. New York: Oxford University Press.

Putnam, H. (1975). Mind, language, and reality. New York: Cambridge University Press.

Rehder, B. (1999). A causal model theory of categorization. In The Twenty-First Annual Conference of the Cognitive Science Society (pp. 595-600). Vancouver, British Columbia.

Ricciutti, H.N. (1965). Object grouping and selective ordering behaviors in infants 12 to 24 months old. Merrill-Palmer, 11, 129-148.

Rips, L. (1989). Similarity, typicality, and categorization. In S. Vosniadou and A. Ortony (Eds.), Similarity and analogical reasoning. New York: Cambridge University Press.

Shanks, D. R., & Dickinson, A. (1987). Associative accounts of causality judgment. In G. H. Bower (Ed.), The psychology of learning and motivation: Advances in research and theory, Vol 21 (pp. 229-261). San Diego, CA: Academic Press.

Smith, L B.; Jones, S. S.; Landau, B.. (1996). Naming in young children: A dumb attentional mechanism? Cognition, 60, 2):143-171.

Soja, N., Carey, S., & Spelke, E.S. (1991). Ontological categories guide young children's inductions of word meaning: object terms and substance terms. Cognition, 38, 179-211.

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. Psychological Review, 99, 605-632.

Spirtes, P., Glymour, C. & Scheines, R. (1993, 2000). Causation, prediction, and search. Cambridge, MA: MIT Press.

Starkey, D. (1981). The origins of concept formation: object sorting and object preference in early infancy. Child Development, 52, 489-497.

Strevens, M. (2000). The essentialist aspect of naive theories. Cognition, v74 (n2):149-175.

Sugarman, S. (1983). Children's early thought: developments in classification. New York: Cambridge University Press.

Tardiff T,; Gelman, S. A.; Xu, F. (1999). Putting the "noun bias" in context: A comparison of English and Mandarin. Child Development, 1999 May-Jun, v70 (n3):620-635.

Tolman, E. C. (1932). Purposive behavior in animals and men. New York: The Century Co.

O'Keefe, J., & Nadel, L. (1978). The hippocampus as a cognitive map. New York: Oxford University Press.

Watson, J. S.; Ramey, C. T. Reactions to response-contingent stimulation in early infancy. In Cognitive development in infancy. John Oates, Ed; Sue Sheldon, Ed; et al. Lawrence Erlbaum Associates, Inc, Hove, England UK. 1987.

Waxman, S.R., & Markow, D.B. (1995). Words as invitations to form categories: Evidence from 12- to 13-month-old infants. Cognitive Psychology, 29, 257-302.

Wellman, H.M. (1990). The child's theory of mind. Cambridge, MA: MIT Press.

Wellman, H.M. & Gelman, S.A. (1997). Cognitive development: foundational theories of core domains. Annual Review of Psychology, 43, 337-375.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. Cognition, v69 (n1):1-34.

Woodward, A.L., Markman, E.M., & Fitzsimmons, C.M. (1994). Rapid word learning in 13- and 18-month-olds. Developmental Psychology, 30, 553-566.

Z

X Y

A. Parties are a common cause of wine-drinking and insomnia

Z X Y

B. Parties cause wine-drinking which causes insomnia

Fig. 1. Two alternate causal structures for three variables.

ZSXWRY

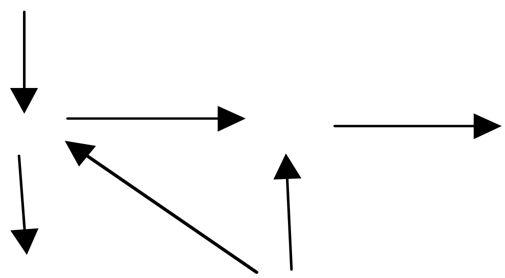

Fig. 2. A causal graph

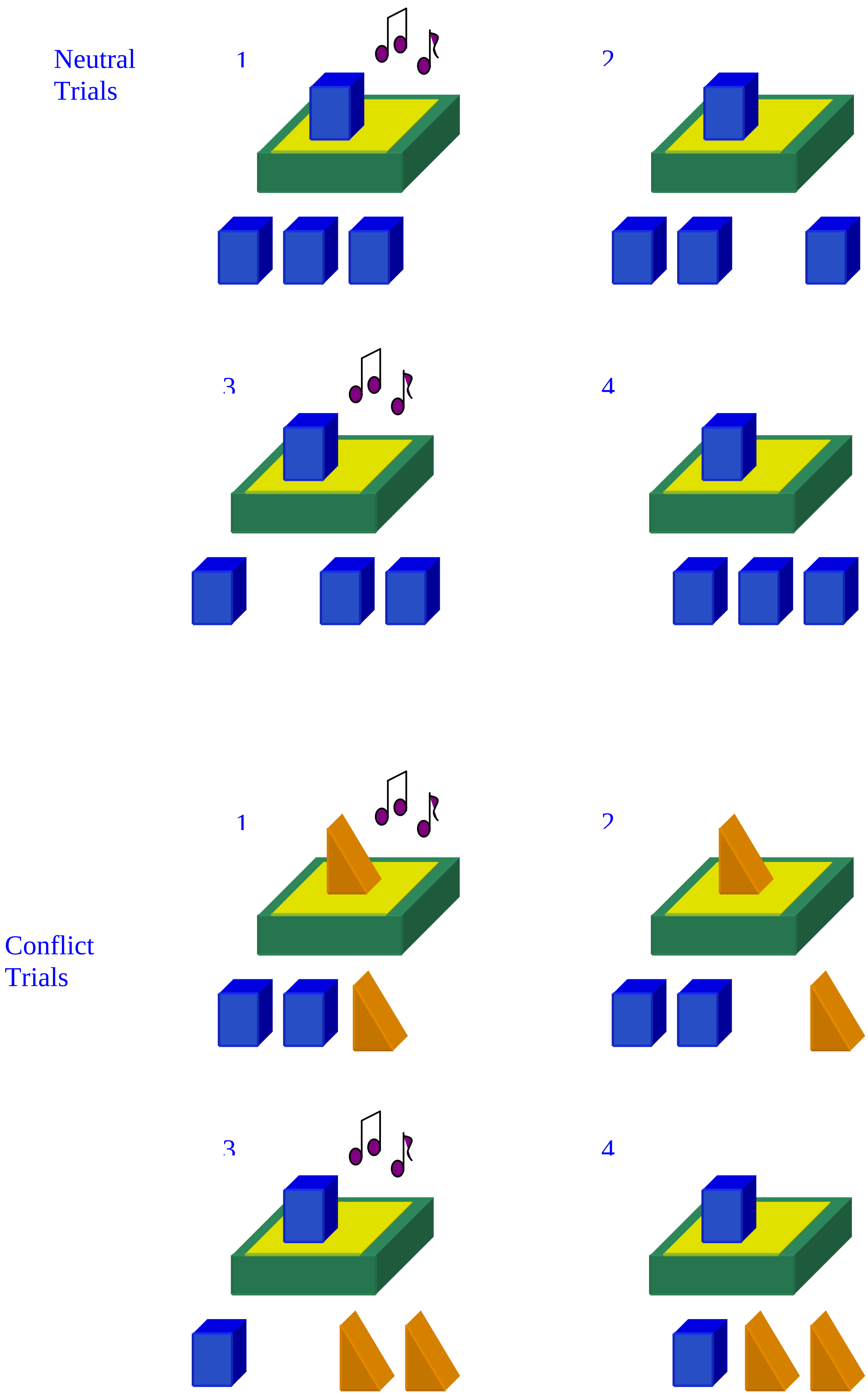

Fig. 3. The blicket detector – Neutral and conflict trials