Developmental Psychology

Learning About Causes From People: Observational Causal Learning in 24-Month-Old Infants

Andrew N. Meltzoff, Anna Waismeyer, and Alison Gopnik

Online First Publication, February 27, 2012. doi: 10.1037/a0027440

Developmental Psychology

0012-1649/12/$12.00 DOI: 10.1037/a0027440

Learning About Causes From People: Observational Causal Learning in 24-Month-Old Infants

Andrew N. Meltzoff
University of Washington

Anna Waismeyer and Alison Gopnik
University of California at Berkeley

How do infants and young children learn about the causal structure of the world around them? In 4 experiments we investigate whether young children initially give special weight to the outcomes of goal-directed interventions they see others perform and use this to distinguish correlations from genuine causal relations—*observational causal learning*. In a new 2-choice procedure, 2- to 4-year-old children saw 2 identical objects (potential causes). Activation of 1 but not the other triggered a spatially remote effect. Children systematically intervened on the causal object and predictively looked to the effect. Results fell to chance when the cause and effect were temporally reversed, so that the events were merely associated but not causally related. The youngest children (24- to 36-month-olds) were more likely to make causal inferences when covariations were the outcome of human interventions than when they were not. Observational causal learning may be a fundamental learning mechanism that enables infants to abstract the causal structure of the world.

Keywords: causal learning, imitation, social learning, action representation, predictive looking

How do children learn about the causal structure of the world around them? One way they might learn is by watching what happens when other people do things. This kind of observation allows people to learn about many everyday tools and skills. In arenas from cooking to hunting to car mechanics to child rearing, we watch what ensues when other people act and use that information to figure out how things work and how to act ourselves. For example, we might see that when a gardener hits a tree with a stick the marauding raccoon runs off, that when a caregiver spins a crib mobile the baby stops fussing, or that when someone flicks the switch on the wall the light goes on. When we observe these actions and their consequences, we might infer the causal relations between sudden sounds and intimidated raccoons, spinning mobiles and happy babies, and switches and lights, and put this information to use ourselves. This kind of learning—which we will call *observational causal learning*—plays a particularly crucial role in the informal apprenticeships that have been the primary teaching method for most people through most of history, long before formal education became prevalent (Meltzoff, Kuhl, Movellan, & Sejnowski, 2009; Rogoff, Paradise, Arauz, Correa-Chávez, & Angelillo, 2003; Tomasello, 1999).

Observational causal learning has advantages over other kinds of causal learning that have been described in the literature. Traditionally, philosophers and developmental psychologists have focused on three forms of causal learning.

First, children might use specific, narrowly tuned spatiotemporal parameters and movement patterns as cues to causality, such as the patterns of contact and launching that ensue when one ball collides with another (Michotte, 1962). Infants do indeed seem to be sensitive to such cues (Kotovsky & Baillargeon, 1994; Leslie, 1984b; Leslie & Keeble, 1987; Oakes & Cohen, 1990; Scholl & Tremoulet, 2000).

Second, children might learn the relations between their own willed actions and the immediate effects of those actions, as Piaget (1954) suggested. When children act on the world, they may assume that their action causes a change in objects. This can be seen in infants' early contingency learning (Rovee-Collier, 1987; Watson & Ramey, 1987). For example, very young infants will quickly learn to move a particular limb to activate a mobile. Piaget originally suggested that, in some primitive sense, these infants might understand that kicking one's right foot causes the mobile to jiggle. There is evidence from imitation as well as looking time measures that infants might make similar inferences when they see other people act, not just when they act themselves. That is, they might infer that when someone else intentionally and directly acts on an object, that action causes a change in that object. Several different types of studies suggest that infants may understand simple actions on objects in this way (Leslie, 1984a; Meltzoff,

Andrew N. Meltzoff, Institute for Learning & Brain Sciences, University of Washington; Alison Gopnik and Anna Waismeyer, Department of Psychology, University of California at Berkeley.

This research was supported in part by grants from the James S. McDonnell Foundation (Causal Learning Network), National Institutes of Health (HD-22514), and National Science Foundation (OMA-0835854). The views expressed are the authors' and do not necessarily represent those of the National Institutes of Health or National Science Foundation. Portions of this work were presented at meetings of the Cognitive Development Society, October 2009, in San Antonio, Texas, and the International Conference on Infant Studies, March 2010, in Baltimore, Maryland. We thank Elizabeth Bonawitz, Laura Schulz, Jim Woodward, and David Danks for helpful discussions; Emily Blumenthal, Lucia Jacobs, and Amy Cook for collaboration on early phases of the work; Craig Harris for assisting in child testing and data analysis; and Calle Fisher and Rechele Brooks for help.

Correspondence concerning this article should be addressed to Andrew N. Meltzoff, Institute for Learning & Brain Sciences, University of Washington, Box 357920, Seattle, WA 98195. E-mail: meltzoff@u.washington.edu

1995, 2007a, 2007b; Muentener & Carey, 2010; Saxe, Tenenbaum, & Carey, 2005; Saxe, Tzelnic, & Carey, 2007). Taken together, this research shows that young infants imitate actions on objects and seem to expect that direct contact between agents and objects will lead to changes in those objects as measured by looking time.

However, children who relied solely on these first two processes could only learn a limited set of causal relations. Michottean causation only applies to a very narrow set of physical cases—there is much more to causality than billiard ball collisions. As Piaget (1930) himself pointed out, Piagetian agent-based causation is limited to causal relations between the infant's actions and their immediate outcomes and does not extend to causal relations among events in the external world. Indeed for this reason, Piaget thought that young children were broadly "precausal" (Piaget, 1930). Children might understand how actions (whether their own or others') cause a stick to bang, but still fail to understand the relation between the banging stick and other events that causally follow from it "downstream."

Relationships between loud noises and raccoon intimidation, mobiles and baby distraction, and switches and lights, for example, do not involve contact and launching (Michotte), nor are they direct causal relations between actions and the immediate effect of those actions (Piaget). Nonetheless, they are causal.

A third idea is that children could learn causal structure by simply noting the correlations and associations in the events in the world around them (e.g., Rogers & McLelland, 2004). Even young infants are sensitive to patterns of statistical covariation and will associate some events with others (e.g., Kuhl, 2004; Saffran, Aslin, & Newport, 1996). However, just detecting covariations or associating events is inadequate for abstracting causal knowledge. Often, one event will follow another without being cause and effect. For example, lung cancer is correlated with having tobacco-stained yellow fingers, but yellow fingers do not cause cancer. Given all the covariations in the world, how do children know which of the systematic covariations they detect to treat as causal? Children would have to use some other information to decide when they are dealing with causes and when they are dealing with mere correlations. If Michottean effects and Piagetian agency learning are too narrow, association is too broad.

Recently, a number of philosophers, psychologists, and computer scientists have suggested a newer account of causal knowledge and learning, called the "interventionist" account (e.g., Gopnik & Schulz, 2007; Woodward, 2003, 2007). On this view, knowing that X causes Y means knowing that if one intervened, that is, if one acted to change X, Y would also change. Other things being equal, if one bangs the stick, one will influence the raccoons. This view helps to distinguish causal relations from mere correlations. Correlations or associations that are not causal do not support interventions in this way. The bark of the tree, for example, might flake when one hits the tree with the stick. As a result, the flaking might be correlated or associated with banging the stick, and therefore with the flight of the raccoons, but deliberately flaking the bark would leave the raccoons unfazed. We would say that the banging but not the flaking caused the raccoons to leave. This is why, in science, experimental interventions provide more powerful information about causality than simple correlations do. When we experimentally intervene to change the rate of cigarette smoking, we see a change in cancer rates. We would not get this result if we intervened to get people to wash their fingers. The interventionist view of causation has become increasingly influential in both philosophy and computer science (e.g., Pearl, 2000; Spirtes, Glymour, & Scheines, 1993).

On this interventionist account, then, the outcomes of interventions on the world might be a particularly powerful source of causal information, in everyday life as well as in science. Such inferences would allow a wider range of causal knowledge than Michottean cues or Piagetian agency learning, and they would allow a more appropriately restricted set of inferences than simple association. Children might infer that when they themselves consistently act to change object X, and object Y changes, X and Y are causally related. These inferences would go beyond the Piagetian inference that acting on object X will lead X to change. Research suggests that 4-year-old children do indeed learn about causal structure by experimenting in this way (Schulz & Bonawitz, 2007). When these children see a "confounded" correlation—for example that pressing two levers simultaneously leads two toys to pop up—they will experiment in a way that enables them to figure out the causal relations. They will press each of the levers in turn and observe the effect on the toys.

However, making similar causal inferences based on seeing the interventions of others—what we are calling observational causal learning—might be even more helpful, because adults and older siblings engage in a wider range of causal interactions with people and things than the young children themselves. Watching the further outcomes of the interventions of others could tell one which relationships among objects and events are most relevant for one to investigate through one's own interventions. Moreover, paying special attention to those outcomes could direct one to just the causal relationships that are most important to master in one's particular culture.

We call this observational causal learning, then, to distinguish it from other types of learning in the literature. It is observational because it involves children watching what other people do rather than acting themselves and observing the outcomes of their actions (as in operant conditioning and trial-and-error learning). It is causal because it allows children to learn a wide range of causal relations and is not limited, for example, only to learning particular spatiotemporal relationships like contact and launching. It is also causal because it supports new interventions and actions on the world, rather than simply allowing children to track associations and correlations. Observational causal learning is a powerful mechanism for inducing rapid learning of causal relations.

When do children begin to be capable of this kind of observational causal learning? There are two conceptual issues embedded in this question. First, how can one tell whether children have made a genuinely causal inference, rather than some other kind of inference? Second, how does one know that that inference was the result of observational learning, rather than some other kind of learning?

The interventionist account suggests an answer to the first question. If children genuinely think that X causes Y, they should act on X in order to bring about a change in Y. This kind of goal-directed action would go beyond simple forms of imitation. We know that infants will imitate novel actions on objects: If a 14-month-old sees an experimenter touch his head to a machine, the child will do likewise (Meltzoff, 1988). This, however, does not address the question of whether children recognize the downstream causal relations between events in the world that follow actions. For example, do infants understand that touching one machine might cause another machine to activate—that the action causes X, which causes Y? Or do they just want to imitate head touching?

There is some recent evidence that 2-year-old children who see an action on X consistently lead to a change in Y will both imitate the action on X and look toward Y (Bonawitz et al., 2010; Meltzoff & Blumenthal, 2007). Again, however, we do not know whether this indicates genuine causal understanding. Children might imitate the action on X, and then simply associate that action with the change in Y they have observed before. (As we might predict that yellow fingers will be associated with cancer.) They might produce the action X for its own sake and then expect that Y will change, without producing the action in order to make Y change.

In fact, the literature on "overimitation" suggests that children may sometimes reproduce unnecessary details of the adult action that do not appear to be causally relevant to outcomes. These children do indeed seem to be imitating the actions for their own sake rather than to bring about a result. These findings also suggest that, by itself, imitation of actions on objects need not be an index of causal knowledge about how the objects work (Horner & Whiten, 2005; Lyons, Young, & Keil, 2007; McGuigan, Whiten, Flynn, & Horner, 2007; Nielsen & Tomaselli, 2010). On the other hand, we know that children are not limited to rote imitation of observed actions (Williamson, Jaswal, & Meltzoff, 2010; Williamson, Meltzoff, & Markman, 2008); and they can sometimes reenact the goals of the actor, bringing about the same causal outcome rather than reproducing the surface actions themselves (Buchsbaum, Gopnik, Griffiths, & Shaffo, 2011; Johnson, Booth, & O'Hearn, 2001; Meltzoff, 1995; Phillips & Wellman, 2005).

How can one test whether very young children have used observational learning to infer a causal relation between two events rather than simply associating the events or imitating the actions that they see? If children think that X causes Y, but Z does not, they should choose to act on X rather than Z to bring about Y. If a child sees an adult perform actions on two objects, one that leads to an effect and one that does not, imitation should lead the child to be equally likely to imitate either action. But if the child understands the causal relations between the objects and wants to bring about the effect, the child should choose to act only on the object that actually caused that outcome.

Using this logic, we designed a new *causal two-choice procedure*. Infants saw the experimenter perform two actions on two objects equally often. One was consistently followed by an effect, and the other was not. Then the infants were given a chance to produce the effect themselves. Could the infants go beyond action imitation per se and use the causal relations they had learned to choose the causally efficacious object? Would they choose to act on that object in order to bring about the effect and leave the other object alone?

Of course, the effect might just make one action more salient than another and so more likely to be imitated. To further test whether the inferences were causal, we used an additional and perhaps even stronger index of genuine causal understanding, namely, sensitivity to temporal order—the cause must always precede the effect. To test whether the children were making causal inferences, we introduced a control condition in which the temporal order of the interventions and outcomes was reversed. Now the effect preceded the adult's actions and the resulting potential cause, rather than vice versa, so that the effect could not be caused by the outcome of the action. If children in the experimental condition were making a genuinely causal inference, the results should fall to chance, and they should not choose to act on one or the other object, although everything other than timing remains the same.

This still leaves the question about the special importance, if any, of observing interventions brought about by people. There is evidence, particularly in the "blicket detector" paradigm, that 2- to 4-year-old children can learn new causal relations between objects in contexts that involve human interventions (Gopnik et al., 2004; Gopnik, Sobel, Schulz, & Glymour, 2001; Schulz, Gopnik, & Glymour, 2007; Sobel, Tenenbaum, & Gopnik, 2004). When children see a causal relationship between X and Y, but not Z and Y, they selectively act on X, and not Z, to change Y. In fact, 4-year-olds can infer more complex causal relations and make these inferences even when there is no physical contact between the cause and the effect (Gopnik et al. 2004; Kushnir & Gopnik, 2007; Schulz et al., 2007). One study suggests that 18- and 24-month-olds can perform similarly on the simplest of these tasks (Sobel & Kirkham, 2006).

In all these blicket detector studies, children saw the outcomes of the experimenter's interventions and made genuinely causal inferences. They may have been engaging in observational causal learning. However, in these studies children also had considerable additional information that may have triggered a causal inference. In particular, children not only watched the events but also heard causal language—the adult described the unfolding events in causal terms. This may have provided an important cue to trigger causal inference: If the same language is used to describe the ongoing event shown by the adult and then to ask the children to act themselves, the language may provide glue between the observed and to-be-executed intervention (Bonawitz et al., 2010). Also, the simple blicket detector studies, those using 18- to 30-month-olds (Gopnik et al., 2001; Nazzi & Gopnik, 2000; Sobel & Kirkham, 2006), involved direct physical contact between the cause and the effect, and that also may have been a cue to causation (Bonawitz et al., 2010). Making direct contact between the block and the blicket detector leads to the change in the detector, and this spatial and temporal conjunction may support, or be necessary for, the children's causal representation. Moreover, since these studies all involved human interventions and their outcomes, we do not know whether observing the pattern of covariation alone was enough to trigger causal inferences, or whether there was some special advantage to the fact that these events involved the intentional actions of other people.

In short, these earlier experiments involved a wide range of potential cues that the events were causal, including the covariation itself, the causal language, the direct contact between cause and effect, and the fact that the covariations were the outcome of human interventions.

In the current studies we isolated the effects of the interventions: Children did not hear causal language, the effects and the potential causes were spatially separated, and we compared covariations that did and did not result from human interventions. In particular, we compared the children's performance in the intervention condition with a "natural covariation" condition. In this condition children saw the same correlations between events and outcomes, but the causal events were not the result of human actions—instead they unfolded with no human involvement. We also varied the objects, events, and the nature and complexity of the actions that were required across four experiments in order to increase the generalizability of the results.

Experiment 1

The experiment involved two independent groups. In the human intervention group, children first saw the experimenter act on two boxes to make each box light up and make noise (the cause). When one box was lit, another machine, a marble dispenser, produced a marble (the effect); when the other box was lit, the marble dispenser did nothing. Then children learned how to produce the cause themselves (but without observing the effect)—they learned how to make the boxes light up and make noise. Finally, children were prompted to produce the effect themselves—we asked them to get a marble. In the natural covariation group, the children saw the same sequence of events, but in the first phase the boxes simply activated spontaneously with no human intervention. The primary dependent measure was whether the child chose to intervene on the box he or she had seen make the marble dispenser go in order to get the marble.

Method

Participants. The participants were 47 children between 24 and 34 months old (M = 27.87 months, SD = 2.28). Children were randomly assigned to one of two independent groups: the human intervention group (n = 24; 12 female, 12 male) and the natural covariation group (n = 23; 12 female, 11 male). Children were recruited at infant and toddler centers associated with the University of California at Berkeley after obtaining informed consent from a parent. The sample was primarily middle to upper middle class based on previous analysis of the children from the centers. Additional children were excluded from the final sample because of experimenter error (four), sound sensitivity (two), and unwillingness to participate (two).

Test environment and stimuli. The study took place in testing rooms at the infant and toddler centers. Children sat at a small rectangular table next to the experimenter, with their parent or another familiar caregiver present to ensure that these young children felt comfortable with the setting. A single video camera recorded each session, focusing on the child's upper body and torso and the table with the experimental stimuli.

The stimuli were three machines (see Figure 1). Two of them (the activators) were identical-sized boxes (12.5 cm × 15 cm × 6 cm), but they were visually and auditorily distinguishable. One was decorated with stripes and the other with polka dots. When the boxes were activated, the top panel (made of translucent plastic) lit up, and the box played a sound (one box emitted a typewriter sound and the other a futuristic blooping sound). A third, larger machine, the marble dispenser, was a large white dome-shaped object (41 cm × 26 cm × 22 cm) that was rigged to dispense marbles from a slot on the side facing the child. All three machines were located on a T-shaped wooden surface on top of the table. The two activators were located on the wings, and the dispenser was located on the central leg directly between and adjacent to the two activators. When the child was seated at the chair, the three machines were just out of reach on the table.

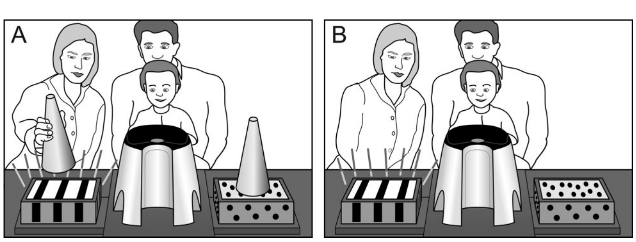

Figure 1. The Phase 1 setup for the human intervention group (A) and natural covariation group (B).

Design and procedure. Before entering the room, children were told that they were going to play the marble game and that the machines would make noise. Children were randomly assigned to either the human intervention or the natural covariation group. The procedure for each group involved three phases.

Phase 1: Observation. Phase 1 for both groups involved observation only.

Human intervention group. When the child entered the room, the two activators had cardboard cones (base diameter = 7 cm, height = 18 cm) on top of them. Children watched as the experimenter, seated next to them, reached over and lifted a cone off the top of each activator (see Figure 1A). The activators were designed with a pressure-sensitive top so that as soon as the experimenter lifted up the cone, the activator lit up and made a noise. When one of the activators (the cause) was activated, it immediately lit up and made its sound, and a hidden confederate immediately triggered the central marble dispenser. (Pilot studies with adults verified that this pattern of activation was perceived to be causal—adults said that the box had made the dispenser go.) When the cone was taken off the other activator, it also lit up and made its sound (possible cause), but this time there was no effect, no triggering of the marble dispenser.

The pattern of covariation presented to the children was such that there were five events in which the effect occurred and five events in which it did not. All children saw the same predetermined pattern of covariation between effective and ineffective actions (ABBABBABAA, where "A" denotes the effective action). The experimenter did not narrate or use causal language to describe the events (e.g., "This box made it go"). She used general phrases to bring the child's attention to the display (e.g., "Let's watch," "Let's look in front," "Did we get a marble?"). Which activator was followed by the marble dispensation, what side it was on, and what sound it played were all counterbalanced across participants within each group.

Natural covariation group. The procedure in the natural covariation group was identical to that used for the foregoing group except for the crucial difference that the cause appeared to happen spontaneously, without human intervention of removing the cones (see Figure 1B). The same verbal phrases were used to bring the child's attention to the display. All three machines were controlled by a confederate hidden behind a one-way mirror. The children watched from their seats with the first experimenter sitting at their side. One of the boxes lit up and made noise, and the dispenser immediately produced a marble. Then the other box lit up and made noise, and the dispenser did not produce the marble. The predetermined pattern of covariation was the same as in the human intervention group (ABBABBABAA).

Phase 2: Practice. Phase 2 was identical for both groups. At the start of Phase 2, the experimenter said, "Now I'm going to show you something different. I'm going to unplug the big machine and put it over here." The experimenter then removed the marble dispenser and placed it out of sight behind the child. From the side of the table opposite that of the child, the experimenter placed one of the cones on each activator. She then demonstrated how to make each of the activators make sounds and light up by lifting the cone off each machine one or two times. The child was encouraged to play with the cones and activators in order to practice making the activator light up and the sounds come on. To make certain that children would be motivated to get the marble in the test phase, rather than simply wanting to play with the activators, Phase 2 was subject controlled in the sense that each child was allowed to practice with the activators until the child no longer expressed interest. Phase 2 ensured that children in both groups were familiar with the cones and could use them to make the activator boxes light up and make sound, and that they had seen the experimenter perform the actions on both boxes.

Phase 3: Test. Phase 3 was also identical for both groups. In Phase 3, to test whether the child performed the correct intervention, the marble dispenser was again placed in the center, between the boxes with the cones on top. The board with the three machines on it was pushed toward the child, and the child was encouraged to "get the marble" while the experimenter sat on the child's side of the table. If the child lifted both cones at once, the marble did not dispense. This was meant to encourage children to continue responding until they made a single choice. If the child did not act immediately, the experimenter encouraged the child to try to get the marble. Finally, if the child continued to refuse to act, the experimenter asked the child a forced-choice question. To do this, the experimenter moved to the opposite side of the table from the child and, while gesturing toward both activators, simultaneously asked, "Which one of these gets you the marble?" Only one child still did not make a choice at this point and was excluded from the analysis.

Scoring. The primary dependent measure concerned the box on which the child first intervened—that is, did the child lift the cone off the correct box? Children's test periods were scored by the experimenter, and 80% were rescored by an independent coder who remained uninformed about the child's test group. The scorers recorded which of the two machines children activated first during Phase 3. The scorers also identified whether the child lifted the cone spontaneously or had to be prompted by being asked the forced-choice question. There were no interscorer disagreements, yielding a kappa of 1.0.

Results and Discussion

Toddlers systematically chose to act on the correct activator to make the marble dispenser go in the human intervention but not in the natural covariation group. The number of toddlers choosing to intervene on the correct activator machine in the human intervention condition (20 of 24 children) was significantly greater than that in the natural covariation condition (12 of 23 children), $\chi^2(1, N = 47) = 5.25$, $p = .02$, $\phi = .33$. The number who did so in the human intervention group was also significantly greater than would have been expected by chance (binomial test, $p = .001$, $g = .33$). If we analyze the spontaneous responses alone, dropping the children who needed forced-choice prompting, the results remain the same. With the spontaneous responses alone, 18 of 22 children chose correctly in the human intervention group, exceeding chance (binomial test, $p = .002$, $g = .32$), and only six of 12 children did so in the natural covariation group.

The human intervention group also acted as a kind of control for the natural covariation group, because the two were identical, save for the human intervention factor. That is, the children's chance performance in the natural covariation group could not simply be the result of the length or complexity of the task, the delay between the observation phase and the test phase, or other, more superficial features that were shared across groups. In particular, the children in both groups had seen the same number of positive and negative associations between the boxes and the dispenser in Phase 1.

Of course, in Phase 1 the children saw more human activity on the boxes in the one group than the other. One might wonder whether this might have led the children in the human intervention group to act more on the boxes in Phase 2, and so perhaps to be more willing and able to act in Phase 3. Similarly, in Phase 1 in both groups, one box was followed by an effect and the other was not, and this might have made that box more salient and so more likely to be acted on in Phase 2. We checked this by rescoring the videos to determine how often the children lifted the cones in Phase 2. The results showed that the children in both groups acted similarly on both boxes in Phase 2, and all the children learned how to activate the boxes. There was no significant difference in the number of times the cones were lifted and the boxes activated in Phase 2 in the natural covariation versus the human intervention group, $t(42) = -0.99$, $p = .33$. In the natural covariation group, children activated the causal box an average of 3.40 times ($SD = 2.28$) and the noncausal box an average of 3.35 times ($SD = 1.57$), $t(19) = 0.15$, $p = .88$. In the human intervention group, children activated the causal box an average of 3.88 times ($SD = 1.90$) and the noncausal box an average of 3.96 times ($SD = 1.90$), $t(23) = -0.34$, $p = .74$. (In the natural covariation group three families declined video consent, so there are 20 participants in the analysis of that group and 24 in the human intervention group.) Thus the boxes were equally attractive to the children and were manipulated about equally in Phase 2, but when children were posed the causal problem of making the marble dispenser work in Phase 3, children in the human intervention group selectively intervened significantly more often on the causal object. There was no significant difference in performance as a function of age (median split for the 10-month age range tested) in the number of children who intervened on the correct activator, either for the entire sample ($p > .12$) or if we broke this down and considered each of the experimental groups separately.

Experiment 2

The toddlers in Experiment 1 were remarkably good at learning the causal relation between the boxes and the marble dispenser in the human intervention group, in spite of the fact that there was no guidance from causal language during the demonstration period nor any physical contact between the cause and the effect. They used the causal relation to plan a new action themselves in the test phase. The two-choice procedure goes beyond earlier findings showing that toddlers can imitate actions on objects and anticipate effects. However, the children in our study did not learn the causal relationship in the otherwise similar natural covariation group.

This latter finding echoes the results of Bonawitz et al. (2010), who reported that 2-year-olds were unable to use natural covariation to infer a causal relation and act appropriately unless causal language or contact was involved. In contrast, 4-year-olds in the same study were apparently able to learn a causal relation based on natural covariation alone. The children in the Bonawitz et al. study saw a single movement—one cube spontaneously moving to hit another cube, which was tethered by a wire to a toy plane (the plane's propeller spun when the contact occurred). The measure was whether the children reproduced the block movement and looked to the toy. The 4-year-old children in the natural covariation condition might have been primarily driven to reproduce the single action they saw rather than intervening in order to bring about the outcome. Would they be able to learn from natural covariation with the current, more stringent two-choice procedure?

To determine whether these older children could, in fact, make this sort of genuinely causal inference from natural covariation, we tested 3- and 4-year-old children using the same two-choice procedure developed in Experiment 1. Given that our sample of 2- to 3-year-olds in Experiment 1 were already near ceiling in the human intervention group (20 of 24 succeeded), the 3- and 4-year-olds used in Experiment 2 were only tested in the natural covariation condition to check for developmental change in that condition.

Method

Participants. The participants were 70 children, 34 three-year-old children (M = 3.48 years, SD = 3.36 months) and 36 four-year-old children (M = 4.50 years, SD = 3.62 months). Children were tested at preschools associated with the University of California at Berkeley after obtaining informed consent from the parents. The sample was primarily middle to upper middle class based on previous analysis of the school demographic data. Half the 4-year-old participants and 41% of the 3-year-old participants were female. Additional children were excluded from the final sample because of experimenter error (eight), sound sensitivity (two), and unwillingness to participate (three).

Stimuli, procedure, and scoring. The stimuli, procedures, and scoring were identical to those in Experiment 1 except that the caregiver was not present (the testing took place in a preschool after parents dropped off their children, and these older children were comfortable without the caregivers' presence). Testing took place in a quiet research room at the preschool. There were no interscorer disagreements, and thus there was a kappa of 1.0.

Results and Discussion

Collapsed across the two age groups, 45 of 70 children chose to intervene on the correct activator, which is significantly greater than would have been expected by chance (binomial test, p = .01, g = .14). This overall effect can be broken down by age. The number of 4-year-olds choosing to intervene on the correct activator machine (24 of 36) exceeded chance levels (binomial test, p = .03, g = .17), but the number of 3-year-olds who chose to intervene did not (21 of 34; binomial test, p = .11). There was no significant difference at either age (or collapsed across age) in the proportion of children who chose to intervene on the correct machine spontaneously versus after a forced-choice (Fisher exact tests: 3-year-olds, p = .25; 4-year-olds, p = .65; collapsed overall: p = .55).

At least by 4 years of age, children succeeded in making causal inferences in the natural covariation group, replicating the basic result in Bonawitz et al. (2010) with a more stringent test. This is a more stringent test, because in the current study there was no causal narrative during the demonstration of the causal event, no physical contact or wire connecting the cause and effect, and the children also had to selectively choose between two actions they had seen the experimenter perform and had performed themselves, ensuring that simple imitation could not have produced the result. The 4-year-olds could use covariations to infer that one box caused the effect but the other box did not, even when those covariations were not the outcome of human actions.

Experiment 3

The results of Experiment 1 suggest that 2-year-olds are adept at using the intervention of others to make causal inferences. These children chose which intervention to make based on the covariation of the events that followed. They did not simply imitate but used the actions and outcomes to make an inference about the causal relation between the machines and the marble dispenser. Experiment 1 also suggested that the 2-year-olds did not make these causal inferences when they saw a similar pattern of covariation between the cause and the effect that was not the outcome of human interventions, although Experiment 2 showed that 4-year-olds did.

In Experiment 3, we sought to add to this knowledge in four ways. First, we employed a different control condition to test more rigorously whether younger children go beyond association to make genuinely causal inferences. The new control involves a temporal reversal: The human agent acts in the same way to bring about the first event (X), but the effect (Y) precedes rather than follows that action. Therefore there could not have been a causal relation between X and Y, although the two events were still associated. Second, we measured predictive looking in addition to the action measure. Here we assessed whether children visually anticipate that the remote effect will occur when they act correctly on the machine. Predictive looking on its own would not provide definitive evidence that children understand the action causally, because children could simply associate the action with the effect. But if children conjointly choose the correct intervention on the two-choice procedure and at the same time anticipatorily look to the effect, it strengthens the argument that children are genuinely reasoning in a causal way. Third, we used two trials rather than one, to provide a more sensitive measure of children's abilities. Fourth, we tested younger infants: a group of 24-month-olds within 1 week of their birthday.

Method

Participants. The participants were 32 twenty-four-month-old infants (M = 23.98 months ± 7 days of their 2nd birthday). Half the participants were female. Infants were recruited by telephone from the University of Washington's computerized participant pool. Preestablished criteria for admission into the study were that the infants be full term, normal birth weight, and have no known developmental concerns. According to parental report, the racial/ethnic makeup of the participants was 81.3% White, 3.1% Asian, 12.5% other (e.g., more than one race), and 3.1% not disclosed, with 3.1% being of Hispanic ethnicity. The sample was primarily middle to upper middle class, based on previous analyses of this university participant list. An additional 14 infants began testing but were excluded due to experimental or equipment error.

Test environment and stimuli. Each infant was tested in the laboratory while seated on his or her parent's lap at a black rectangular table. Two digital cameras recorded the session, each on a separate recorder. The main camera provided a close-up of the infant's face, hands, and upper body; the other focused on the experimenter. A character generator added synchronized time codes (30 per s) onto both digital recordings, which were used for subsequent scoring from the digitized video record.

There were two sets of objects (see Figure 2). They were arranged on the table in a way that was similar to that in Experiments 1 and 2, with the effect in the middle and the two potential causes on either side. Set A consisted of a stick, two button boxes, and a translucent egg-shaped object. When the stick was used to press one of the buttons (cause), it remotely made the egg light up (effect). The two button boxes were identical except that one was black and the other white. The button boxes (16.5 cm × 15.2 cm × 5.5 cm) were tilted 30° off the horizontal; the egg was 8.5 cm tall and 6.5 cm wide. Set B consisted of two hemicircle platforms, a small rubber dog, and a smoke-colored plastic box. When the dog was placed on top of one of the wooden platforms (cause), a red X shape lit up inside the box (effect). The two wooden platforms were identical save that one was painted brown and the other pink (height = 3.8 cm; circle diameter = 13.3 cm). The red X was visible when lit up and formed by LEDs inside of the box (15.8 cm × 15.3 cm × 15.3 cm).

Design: Causal versus control events. The temporal parameters of the two events are shown in Figure 3. The infants were randomly assigned to the causal event ($n = 16$; Figure 3A) and the control event ($n = 16$; Figure 3B), with half the participants in each group being female. Within each group each infant received two trials, using Set A and Set B objects. The design was fully counterbalanced with respect to type of event, sex of child, which set was used on Trial 1, side of first demonstration, and left–right side of the object that was associated with the light. The side of the causal or associated object was changed between trials for each infant: If the left button box was the cause on Trial 1, the right platform was the cause in Trial 2. This ensured that if children simply produced the same response on both trials (e.g., choosing the left object to act on in the first trial and repeating this in the second trial), they would be at chance level.

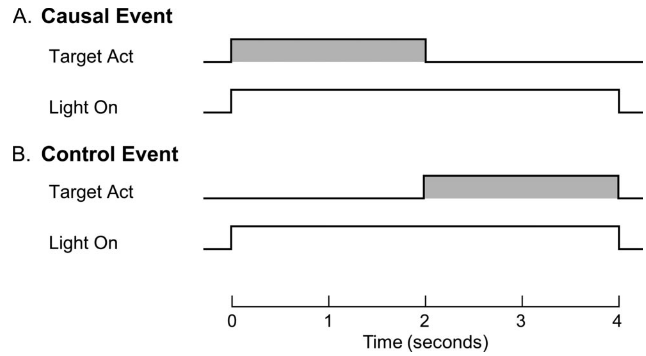

Figure 3. The causal and control events in Experiment 3 contained the same elements. The elements were simply arranged differently in time. (A) For the causal event, the light came on immediately when the button was pushed by the stick. The button pushing appeared to cause the light to come on. (B) For the control event, the light came on 2 s before the button push. This did not give a causal impression. In both events the target act and the light illumination were equally associated, inasmuch as they overlapped for the identical duration (2 s).

Procedure. Upon arrival at the university, families were escorted to a waiting room where they completed consent forms. They were then brought to the test room and seated at a table across from the experimenter, and the experimenter handed the infant an assortment of small toys to acclimate the child. After the infant seemed comfortable, he or she was presented with either the causal or control event.

Causal event. The experimenter brought the test objects from below the table one by one and placed them on the table surface. This helped to emphasize that the objects used for the cause and effect were spatially distinct. The objects were placed out of reach of the child, 2 cm from the adult's side of the table, so that the infant observed the display but could not interact with the test objects. Next the adult demonstrated the events.

For illustrative purposes, the procedure is described with Set A objects; the same procedure is also followed with Set B on a second trial (order counterbalanced). As shown in Figure 4, there were two boxes (candidate causes) differing only in color. When the experimenter pressed the button on one of the boxes, the effect immediately occurred (the egg light in the center came on). This

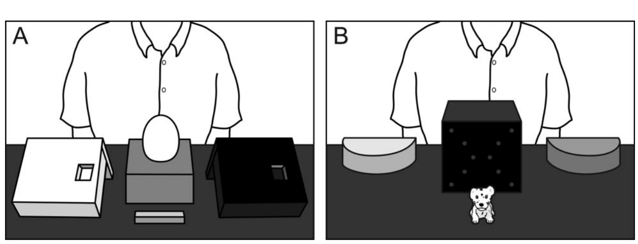

Figure 2. Experiment 3 used two sets of test objects: Set A (button boxes) and Set B (platforms). The stick was used to push the buttons; the dog was put on the platforms. These target acts (causes) made other things happen at a distance (effects): The central object lit up (egg or black box).

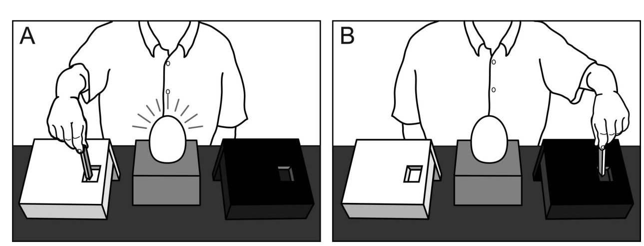

Figure 4. In the demonstration phase in Experiment 3, the experimenter acted on two objects differing only in color. (A) When the button on one of the boxes was pushed, the light came on. (B) When the button on the other box was pushed, it did not.

pattern of covariation provided a compelling impression of causality. Pilot work with adults verified that adult observers thought that pressing the button caused the light to come on. (In fact, the light was activated by the experimenter via a foot pedal, which triggered a radio signal transmission that activated the light.) For ease of description, we say that performing the target act on the activator box "caused" the light to come on, and performing it on the other did not.

The experimenter did not provide a narrative about what he was doing or use causal language (e.g., "The button makes the light go") during the demonstration. The language was confined to bringing the child's attention to the display, saying, "Look!" "It's my turn," and "Here we go, look at this." The experimenter pushed the button on the first box three times during an approximately 20-s period; next, he pressed the button on the other box three times in approximately 20 s. Each time the experimenter pressed the button on one of the boxes (the activator box), it caused the light to come on; pressing the button on the other box in an identical manner had no effect. The adult did not look at the light when he pressed the button, because we did not want to model looking at the effect.

After the demonstration period, the adult gave the child the stick, pushed the two boxes (but not the effect) forward so they were within reach, and told the child, "It's your turn" and "This is for you." A 20-s response period was electronically timed, starting from when the child touched the stick. At the termination of the response period, the test objects were cleared from the table. The same procedure was then repeated for Trial 2 with the second set of test objects.

Control event. The same elements were used, but the timing was changed. For one of the boxes, the light came on before the button was pushed (see Figure 3B). Because the light was already on when the adult pushed the button, it did not look like the button push activated the light. In this condition, the light being lit still overlapped with the human action for 2 s, just as in the causal case. For ease of description, we say that performing the target act was "associated with" the light for one of the boxes but not the other. All other aspects of the protocol for the control event were identical to that of the causal event.

Scoring and dependent measures. The response periods for both experimental groups were identical inasmuch as each infant had two 20-s response periods. To ensure blind scoring of the data, a new video record was made by deleting all the warm-up and demonstration periods. It comprised only the response periods and thus contained no artifactual clue as to the children's test group. It was scored in a random order by a coder who was kept unaware of the test group of the children. Manual actions and looking behavior were both scored.

Manual act score. Each infant had two trials, one with the stick and button box and the other with the dog and platform. For the button box, the target act was using the stick to press the button. The score reflected the box to which infants directed their first target act. A $+1$ indicated they performed the target act on the correct box (the one that caused or was associated with the light coming on), and A -1 indicated they performed the target act on the incorrect box. For the platform, the target act was placing the dog on the platform, and each infant received a score of $+1$ or -1 in the same manner for this task.

For each infant, a total score was calculated by summing the two trials, which constituted the "manual act" score. A score of $+2$ (or -2) indicates that the infant produced a target act on the correct (or incorrect) object on both trials. A score of $+1$ (or -1) indicates that the infant produced a target act on the correct (or incorrect) object on only one trial and did not produce a target act on the other trial. A score of 0 indicates that the infant produced the target act on the correct object on one trial and on the incorrect object on the other trial. Infants who did not produce a manual target act on either trial were dropped from the analysis, but the data were also reanalyzed with these infants included (assigned a score of 0), and the results did not change—all statistically significant effects remained significant, and all nonsignificant effects remained so.

Manual act + predictive looking score. Infants not only responded with manual actions but, in some cases, also coupled this with systematic looking behavior. Predictive looking was scored if the infant immediately looked toward the effect when producing a target act (i.e., pressing a button or placing the dog on a platform). "Immediately" was defined as the child visually fixed the effect while producing the target act or $\leq$1.25 s after it. Such looking was predictive of the effect, because the light did not actually activate synchronously with the child's action (recall that the light was controlled via the experimenter's foot pedal). The light was activated only after the child produced the target act (on either the correct or incorrect box) and looked to the effect so that predictive looking could be scored. Predictive looking was determined through frame-by-frame analysis of the video record. Because each video frame had a time stamp inserted in it, the coder could readily specify when looking, if any, occurred. The frontal camera angle provided a close-up of the child's face, supporting fine-grained scoring of both the manual and looking behaviors (see Scoring Agreement).

On many trials children produced the correct (or incorrect) target act but did not immediately look for the effect, and this was captured by the manual act score (see above). The current dependent measure is a more demanding response and evaluated manual action plus looking. There may be reasons that children produce the target act on the wrong side—they could do so in imitation of the adult, because the adult acted on both sides—but when a child intervenes on an object and immediately shifts head and eyes away from that object to the remote effect before it occurs, it goes beyond action imitation per se and is unlikely to be accidental. The mean latency for predictive looking was 0.43 s after acting, which matches our impression that children were shifting their gaze in order to see the effects of their actions. An advantage of the manual act + predictive look score is that the behavior, though rare, is likely to be informative; a disadvantage is that pilot studies have shown that only a small number of the children produce the manual act while simultaneously making an anticipatory look to the effect.

The manual act + predictive looking score was calculated in the same manner as described above (ranging from $+2$ to -2). For example, if an infant produced correct target acts on both trials and did so with predictive looking both times, the child was assigned a $+2$. If the infant produced the target act to the incorrect object on both trials and did so with predictive looking, the child was assigned a score of -2. Infants who produced no predictive looking on either trial were dropped from analysis, but we also reanalyzed the data with them included (assigned a score of 0), and the results did not change—all significant effects remained so, and all nonsignificant effects remained as such.

Scoring agreement. Scoring agreement was assessed by having a randomly selected 25% of the infants rescored by an independent scorer. There were no intrascorer disagreements on either the manual act or the manual act + predictive looking scores. For the interscorer assessments, there were also no disagreements on the manual act score, and only one for manual act + predictive looking. Cohen's kappa ranged between .85 and 1.00 for all measures.

Results and Discussion

Infants in the causal event group had higher manual act scores (M = 0.75, SD = 1.24) than those in the control group (M = −0.53, SD = 1.06), and the difference between groups was significant, $t(29) = 3.09$, $p = .004$, $d = 1.15$ (the same results were obtained by a nonparametric Mann–Whitney U test, $z = 2.85$, $p = .004$). Moreover, the manual act scores in the causal group significantly differed from zero, $t(15) = 2.42$, $p = .03$, $d = 0.61$, whereas those in the control group did not (the same results obtained by nonparametric Wilcoxon tests). Table 1 provides the distribution of raw scores. For the causal group, the manual act measure was discriminative at the dichotomous level comparing the number of infants showing positive (+1 or + 2) versus negative (−1 or −2) scores (seven vs. one; binomial test, $p = .04$, $g = .38$).

Some infants did more than produce the correct response on the action measure; they also immediately looked toward the effect. Infants in the causal event group had higher manual act + predictive looking scores (M = 1.33, SD = 0.89) than those in the control event group (M = −0.30, SD = 1.06), and the difference between groups was significant, $t(20) = 3.94$, $p = .001$, $d = 1.76$ (Mann–Whitney U test, $z = 2.93$, $p = .003$). Moreover, the scores in the causal group significantly differed from zero, $t(11) = 5.20$, $p < .001$, $d = 1.50$, whereas those in the control group did not (the same results obtained by Wilcoxon tests). Table 1 shows that the manual act + predictive looking measure was highly discriminative at the dichotomous level: In the causal group 11 infants had positive scores compared with one with a negative score (binomial test, $p = .003$, $g = .42$). The predictive looking was not the result of imitating the adult's pattern of behavior, because the adult did not look at the effect when he acted (see Procedure).[1]

Table 1
Experiment 3: Number of Infants as a Function of Experimental Group and Dependent Measure

Group	−2	−1	0	+1	+2
		Manual act			
Causal	1	0	8	0	7
Control	3	4	7	0	1
		Manual act + predictive looking			
Causal	0	1	0	5	6
Control	0	6	2	1	1

Note. Infants not producing a criterion behavior on either trial are not included in the table. The number of infants who did not produce criterion responses in each row are, respectively, 0, 1, 4, 6.

We infer that 24-month-old infants learn a causal relationship from observing the adult's intervention and predict that their own acts will have the same effect on the world as the adult's. Three findings support this. First, infants in the causal group preferentially directed their target acts to the object that caused an effect, even though they saw the adult perform the identical actions on both objects. Second, infants immediately looked to the effect when they acted on the causal object, anticipating the result before it occurred. Third, infants' did not behave this way in the control condition. This shows that the children's success was not the result of superficial factors such as a difference in salience between the two actions due to association with the light. Taken together, the results demonstrate observational causal learning.

Experiment 4

Experiment 3 shows that 24-month-old infants can learn causal interventions from observing the outcomes of the actions of others. However, this does not address the additional question posed earlier in the article: Will children also do this when they see a similar pattern of covariation with no human action at all? Are human actions special?

Experiment 1 suggests that 2-year-old children do not learn causal relations from natural covariation without human intervention, but there were other differences between the two conditions. The Phase 1 demonstration for the human intervention group involved the additional component of manipulating interesting cones that led to the activation of the box. This may have made the human intervention condition more salient or easier to process than the natural condition. Moreover, in Experiment 1 we used a fairly complex causal chain of events. We had to include Phase 2 in order to ensure that children in the natural covariation condition would know how to activate each of the causal boxes. In the current experiment, rather than having to learn to lift a cone off the box to activate it, the children simply had to move an object. The apparatus also allowed us to show the children simple event sequences and vary whether it was or was not produced by human action while holding everything else constant.

To accomplish this, we constructed special objects that could move without (apparent) human intervention. A central disk moved to one side, and when it touched a lateral object, it caused a remote light to come on. When it moved in the other direction and touched a different lateral object, nothing happened (following the two-choice logic). For one group of infants (human intervention group), the adult put his hand on top of the object and slid it laterally to cause the contact with a lateral object; for the other group (natural covariation group), the object moved autonomously

[1] One might wonder whether infants failed to watch the adult's target acts after the "effect" occurred. To check this, we had a coder, who was blind to the experimental hypothesis, review the video records and score whether each infant saw the critical events. For each trial there were nine critical events (three manual acts to one side, three acts to the other side, and three effects). These occurred in each of two trials for 32 participants. Of these 576 events (9 events × 2 trials × 32 infants), there were nine events that infants did not observe (the infant was scored as looking somewhere else when the event occurred). These nine missed events were distributed across six infants, three from Group 1 and three from Group 2. We conclude that infants adequately saw the displays in both groups.

along the same path and touched the object. Infants were randomly assigned to one of these experimental groups. The spatiotemporal properties of the disk movement, the touch, and the remote effect (the light coming on) were identical for the two groups. The key question: Can infants learn the causal intervention and reproduce it themselves merely from observing a natural event? Or is watching human action necessary at this young age?

Method

Participants. The participants were 32 typically developing 24-month-old infants ($M = 23.99$ months ± 7 days of their 2nd birthday). Half were female. Children were recruited in the same manner and with the same inclusion criteria as in Experiment 3. According to parental report, the racial/ethnic makeup of the participants was 71.9% White, 3.1% African American, 21.9% other (e.g., more than one race), and 3.1% not disclosed, with 9.4% of the children of Hispanic ethnicity. Additional infants began testing but were excluded due to equipment or procedural failure (17), fussiness or refusal to watch (four), and parental interference (one).

Test environment and stimuli. The test room, video equipment, and general setup were the same as in Experiment 3. The only difference was that a black platform was situated on the top surface of the table. The platform (83.2 cm × 33 cm × 10.7 cm) covered almost the entire table but left an open section on the back facing the experimenter so that he could manipulate a magnet underneath. The experimenter used a pulley system in the space under the platform to silently control the movements of a magnet that controlled the objects on the top surface. The visual events for the infant unfolded on the top of the platform; the movements below the surface were invisible and silent. To an adult, it looked analogous to a toy train set when a train car spontaneously moves off in one direction; in our case, too, the object moved spontaneously. Pilot work with adults yielded a consistent report, best captured by one participant: "It looks like it moves by itself, and when it touches the block that makes the light come on." No adult guessed that the experimenter was controlling the movements of the disk through hidden pulleys.

Two sets of test objects were constructed (see Figure 5). Set A consisted of the two hemicircle-shaped blocks, a translucent plastic egg (from Experiment 3), and a flat yellow disk. When the disk was moved laterally on the table surface and contacted one of the blocks, the egg lit up. Set B consisted of two wooden bricks (blue and green), a dark plastic box that housed a red X, and a half-moon shape on wheels. When the half-moon-shaped object contacted one of the bricks, the red X came on inside the box. The dimensions of the objects were, for bricks, 14 cm × 7 cm × 3.5 cm; box with red X, 7.7 cm × 7.3 cm × 2.5 cm; flat disk, 5.7 cm in diameter and 1.3 cm high; half-moon-shaped object, 5.1 cm diameter × 2.9 cm × 2.5 cm.

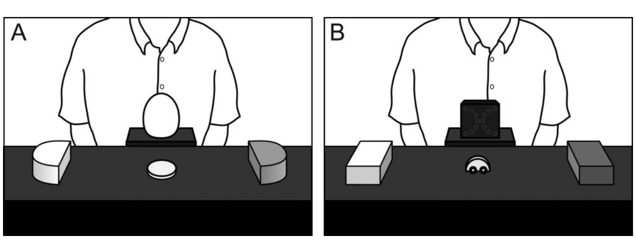

Figure 5. Two sets of objects were used in Experiment 4. The small, central object was either self-mobile (controlled by a magnet beneath the top surface) or moved by a human hand, depending on group assignment. For the human intervention group, the adult's hand moved the object laterally so that it made contact with one of the blocks. This caused the effect (the egg or box lit up). For the natural covariation group, the object moved autonomously to make contact with one of the blocks, which caused the effect. The spatiotemporal trajectory that followed, the timing of the events, and the effect were identical in both groups.

Design and procedure. Infants were randomly assigned to one of two independent groups: the human intervention group ($n = 16$) and the natural covariation group ($n = 16$). Half the participants in each group were female, and the experiment was fully counterbalanced, as described in Experiment 3. After acclimating to the test room, infants were shown one of the experimental demonstrations.

Human intervention. The procedure is described with Set A as the example; the same events unfolded with Set B. For the human intervention group, the adult put his hand on the top of the disk and moved it to one of the blocks. When the disk contacted the block, the remote egg lit up. It looked like the contact caused the light to come on (the light was regulated by a foot pedal and timer, as in Experiment 3). This act was repeated three times, and each time the disk touched the block, the light came on. The disk was then moved to the other block. There was no effect when the disk touched that block. This too was repeated three times. Infants' attention was directed to the events without using causal language by saying, "Look!" "It's my turn," and "Look at this." At the end of the demonstration period, the disk was handed to the child (through sleight of hand, infants were provided an identical disk without a metallic bottom so it would not "stick" to the magnets below the surface). The top surface supporting the blocks (but not the effect) was moved closer to the infant, and as in Experiments 1 and 2, infants were encouraged to play with the objects and produce the effect: "It's your turn, now you make the light go." A 20-s response period was timed. Each infant received two trials, one with Set A and one with Set B (counterbalanced across infants).

Natural covariation. This involved the identical spatiotemporal object movements as the human intervention group. The crucial change was that the disk appeared to move autonomously, without human action. In reality, the disk was controlled by the movement of a magnet below the top surface. The magnet slid silently along a track underneath the surface carrying the disk with it. It was operated through pulleys by the experimenter from below the surface of the table. When the disk moved laterally and touched one of the blocks, the remote effect lit up. This was repeated three times. When the disk moved laterally in the other direction and touched the other block, nothing happened, and this too was repeated three times. The rest of the procedure, event timing, and language used were identical to those of the human intervention group.

Scoring and dependent measures. Both manual behavior and predictive looking were scored as in Experiment 3. The manual target act for this study was pushing the small object (disk or half-moon object) laterally so that it touched one of the side blocks. As in Experiment 3, each infant was assigned a score that ranged from + 2 to −2. The responses were again scored in a random order by a coder who was blind to the infant's experimental group. Scoring agreement was assessed by having a randomly selected 25% of the infants rescored by an independent scorer. For the intrascorer assessments, there were no disagreements on the manual act scores and only one disagreement on manual act + predictive looking. For the interscorer assessments, there were also no disagreements for the manual act scores and one on manual act + predictive looking. Cohen's kappa ranged between .88 and 1.00 for all measures.

Results and Discussion

The manual act score differed as a function of experimental group. Infants in the human intervention group had higher manual act scores ($M = 1.92$, $SD = 0.28$) than those in the natural covariation group ($M = 0.71$, $SD = 1.20$), and the difference between groups was significant, $t(25) = 3.65$, $p = .002$, $d = 1.41$ (Mann–Whitney U test, $z = 3.32$, $p = .001$). Importantly, however, both experimental groups responded systematically. Infants in the human intervention group selectively directed their manual target acts to the causal object more than to the noncausal object, $t(12) = 25.00$, $p < .001$, $d = 6.93$; infants in the natural covariation group also selectively directed their manual target acts to the causal object more than to the noncausal object, $t(13) = 2.22$, $p = .045$, $d = 0.59$ (both also significant by Wilcoxon tests). Table 2 shows that the manual act measure was also discriminative at the dichotomous level: In the human intervention group, 13 infants had positive scores compared with none with negative scores (binomial test, $p < .001$, $g = .50$). In the natural covariation group, nine infants had positive scores compared with two with negative scores (binomial test, $p = .03$, $g = .32$).

As in Experiment 3, some infants not only produced manual target behavior but also conjointly made a predictive look to the effect. Infants in the human intervention group had higher manual act + predictive looking scores ($M = 1.83$, $SD = 0.39$) than those in the natural covariation group ($M = 1.00$, $SD = 0.71$), and this difference between groups was significant, $t(19) = 3.46$, $p = .003$, $d = 1.59$ (Mann–Whitney U test, $z = 2.82$, $p = .005$). Importantly, each group responded systematically and directed more manual act + predictive looking to the causal than to the noncausal object: human intervention group, $t(11) = 16.32$, $p < .001$, $d = 4.70$; and natural covariation group, $t(8) = 4.24$, $p = .003$, $d = 1.41$ (both also significant by Wilcoxon tests). Table 2 shows that the manual act + predictive looking measure was also discriminative at the dichotomous level: In the human intervention group, 12 infants had positive scores compared with none with negative scores (binomial test, $p < .001$, $g = .50$). In the natural covariation group, seven infants had positive scores compared with none with negative scores ($p = .008$, $g = .50$).

Table 2
Experiment 4: Number of Infants as a Function of Experimental Group and Dependent Measure

Group	−2	−1	0	+1	+2
Manual act					
Human intervention	0	0	0	1	12
Natural covariation	1	1	3	5	4
Manual act + predictive looking					
Human intervention	0	0	0	2	10
Natural covariation	0	0	2	5	2

Note. Infants not producing a criterion behavior on either trial are not included in the table. The number of infants who did not produce criterion responses in each row are, respectively, 3, 2, 4, 7.

We reviewed the video records of the infants who did not reach criterion and succeed on the task. Infants were engaged: Some pushed the central object toward the correct side block but insufficiently far to make contact (thus not meeting the response criterion). Infants in the natural covariation group in particular seemed to expect that when they gave the central object a shove, it would move on its own to the target block. In this group, the center object was originally seen to move on its own without human intervention. Speculatively, some infants in this group may have learned the causal links from observation but were not able to convert this into a novel intervention to cause the effect. This is one way in which the natural covariation task is more demanding than the human intervention one.

This experiment shows that infants learn causal relations from observing human interventions, and under these simple and constrained situations some infants can also do so by observing natural events not generated by intentional human action. The results were stronger in the human intervention rather than natural covariation group. Nonetheless, the results in the natural covariation group were statistically significant in their own right, even in this situation that did not involve the support of causal language during the event or spatial contact between the cause and the effect.

General Discussion

Taken together, the results of these four experiments demonstrate both striking capacities and striking limitations in toddlers' causal learning. Two-year-old infants are adept at observational causal learning: They readily learn novel causal relations by observing others act causally. They do not merely learn correlations between observed events, nor do they simply directly imitate the actions of others. Instead they infer a causal link between two events and use this inference to fashion their own interventions in order to achieve the same causal effects they have seen others achieve.

Across three experiments, each involving different actions and events, toddlers went beyond imitation and association and made new genuinely causal inferences about events in the world. They demonstrated this by choosing to intervene selectively on the cause that had been followed by a particular effect. They did this when they were explicitly motivated to act to obtain a marble for an interesting game (in Experiment 1) and when producing the effect was its own reward (Experiments 3 and 4). They did this when they were explicitly prompted to make the effect occur (Experiments 1 and 4) and when they were not (Experiment 3). They also did this when the action that led to the cause was novel (lifting the cones in Experiment 1) and when it was more familiar (moving the disk in Experiment 4), when it involved movement and collision (Experiment 4) and when it did not (Experiments 1 and 3). They also did this in spite of the fact that they did not hear any causal language describing the causal event as it unfolded and there was no spatial contact between the cause and the effect—the effect was a spatially remote independent object in all of the studies. Moreover, infants did not behave in this way when the cause and effect were temporally reversed (Experiment 3), such that the events were merely associated but could not have been causally related.

The results also suggest that toddlers are substantially more likely to make causal inferences when covariations are the outcome of human interventions than when they are not. In Experiment 1, as in Bonawitz et al. (2010), toddlers did not make causal inferences from patterns of natural covariation. In Experiment 4, 24-month-olds made causal inferences from natural covariation, but did much better when the covariations were the outcome of human actions.

An interesting question for further research is why infants proved more adept at the natural covariation condition in Experiment 4 than in Experiment 1 and in the Bonawitz et al. (2010) study. There are several possibilities. First, the causal chain used in Experiment 4 was considerably simpler than that in Experiment 1. If infants could abstract the cause–effect relationship from observing the pattern of natural covariation in Experiment 4, they would already know how to perform the relevant action to generate the cause (i.e., how to slide the disk to make it touch the lateral block). In contrast, even if children could learn the cause–effect relationship from observing the natural covariation in Experiment 1, they still needed to learn the additional information about how to get the box to activate (by lifting the cones) in order to generate the cause in the first place. They had to learn how to produce the cause as well as the cause–effect relationship. Additionally, the spontaneous movement of the object in the natural covariation condition in Experiment 4 may itself have been a cue that led infants to construe the object in a more agent-like way, in contrast to Experiments 1 and 2, where the object did not move.

The current results of the natural covariation condition in Experiment 4 are also stronger than in Bonawitz et al.'s (2010) natural covariation condition, which, like our Experiment 4, did not involve a practice phase and did involve spontaneous movement. One possibility may be that the two-choice procedure helps specify the causal relation for the children (e.g., Kovack-Lesh & Oakes, 2007). Children in the current work see that touching one block causes the effect and that touching the other does not. This might appear to be more complex, but that complexity might actually help the children to make the causal inference correctly. The contrasting case—the juxtaposition between what works and what does not—may help to implicitly "instruct" children, even though it occurs in the context of natural covariation. This suggests that when the world is arranged "just right," infants can learn causal relations from the natural flow of events. Further research will be necessary to unravel these issues, but across all the experiments the difference between human intervention and natural covariation is striking and robust.

Another interesting question concerns developmental change. Using the more complex causal chain in Experiment 2, 4-year-olds, unlike the younger toddlers, were adept at making causal inferences from natural covariation, even with a relatively demanding task. One possibility, as suggested by Bonawitz et al. (2010), is that between 2 and 4 years of age the increasing use of causal language allows children to generalize from the human intervention case to the natural covariation case (cf. Gopnik & Meltzoff, 1986, 1997). The fact that toddlers' causal learning in natural covariation tasks is significantly improved when provided a verbal causal narrative of the unfolding event (Bonawitz et al., 2010) supports this idea. Alternatively, the children may simply accumulate more experience with both types of events and make the generalization that consistent natural covariations support intervention.

Perhaps the most intriguing developmental possibility is that children, at first, use human intervention as a way of constraining the hypotheses they will consider. In the natural world a variety of covariation possibilities are available, and infants may initially focus on those that arise from human interventions. When the possibilities are very tightly constrained by the experimental context, as in Experiment 4, children may be more willing to consider natural covariations as potential cues to causality. In everyday learning situations there might be differences in the kinds of interventions and outcomes that different children see, so that individual children might constrain their hypotheses in different ways. This, in turn, might influence their causal learning.

These findings also have implications for our understanding of children's causal and social learning. As early as 24 months of age, and perhaps earlier, children already have the capacity to infer new causal relationships between a variety of events when those events are the outcome of human actions. This is true even when the events are separated spatially and the causal relation does not involve movement, contact, or launching. This ability clearly goes beyond Michottean perceptual effects and the Piagetian ability to infer that actions cause changes in objects. The children in our studies inferred, for example, that the polka-dotted box would cause a dispenser to produce marbles but the striped box would not, that a button press on the black object but not the white one would cause the remote egg to light, or that using an object to touch the blue brick but not the green one would cause the red X to appear in a box. They used this information to act themselves (and to simultaneously make a predictive look to the effect before it occurred). These infants could infer a wide range of novel causal relations.

Children's inferences are appropriately wide, but they may also be appropriately narrow. Initially weighting one's causal inferences in favor of events that follow an intentional action may help one to avoid spurious correlations. Children's minds may implicitly be applying the maxim that correlation does not necessarily imply causation. Just like the scientists, they may prefer to focus on the outcomes of intentional experiments as a more accurate guide to causal structure. It remains to be seen whether and how these inferences are developmentally related to the early Michottean perceptual effects or to other kinds of causal knowledge.

These results also echo recent findings in the literature emphasizing the importance of social contexts for early learning (e.g., Csibra, 2010; Meltzoff et al., 2009; Tomasello, 1999). In the studies reported here, the children learned best from other people, and this might shape learning in many significant ways. Observational causal learning from people may allow infants to learn which specific causal relations are important in their particular culture or social milieu. In turn this may underlie the informal apprenticeships that are a key feature of teaching in many cultures.

More generally, the fact that very young children are adept at observational causal learning may help explain the rapid development of causal knowledge in the first few years of life. The literature on children's intuitive theory formation shows that before 5 years of age, children have learned about a wide array of everyday causal relationships, including many subtle and surprising ones (e.g., Carey, 2009; Gopnik & Meltzoff, 1997; Gopnik & Wellman, 1994; Wellman & Gelman, 1992). Observational causal learning may be one of the fundamental learning mechanisms that enables children to abstract the causal structure of the world so swiftly and accurately.

References

Bonawitz, E. B., Ferranti, D., Saxe, R., Gopnik, A., Meltzoff, A. N., Woodward, J., & Schulz, L. E. (2010). Just do it? Investigating the gap between prediction and action in toddlers' causal inferences. *Cognition, 115,* 104–117. doi:10.1016/j.cognition.2009.12.001

Buchsbaum, D., Gopnik, A., Griffiths, T. L., & Shafto, P. (2011). Children's imitation of causal action sequences is influenced by statistical and pedagogical evidence. *Cognition, 120,* 331–340. doi:10.1016/j.cognition.2010.12.001

Carey, S. (2009). *The origins of concepts.* New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780195367638.001.0001

Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind & Language, 25,* 141–168. doi:10.1111/j.1468-0017.2009.01384.x

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review, 111,* 3–32. doi:10.1037/0033-295X.111.1.3

Gopnik, A., & Meltzoff, A. N. (1986). Relations between semantic and cognitive development in the one-word stage: The specificity hypothesis. *Child Development, 57,* 1040–1053. doi:10.2307/1130378

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

Gopnik, A., & Schulz, L. (Eds.). (2007). *Causal learning: Psychology, philosophy, and computation.* New York, NY: Oxford University Press.

Gopnik, A., Sobel, D. M., Schulz, L. E., & Glymour, C. (2001). Causal learning mechanisms in very young children: Two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology, 37,* 620–629. doi:10.1037/0012-1649.37.5.620

Gopnik, A., & Wellman, H. M. (1994). The theory theory. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 257–293). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511752902.011

Horner, V., & Whiten, A. (2005). Causal knowledge and imitation/emulation switching in chimpanzees (*Pan troglodytes*) and children (*Homo sapiens*). *Animal Cognition, 8,* 164–181. doi:10.1007/s10071-004-0239-6

Johnson, S. C., Booth, A., & O'Hearn, K. (2001). Inferring the goals of a nonhuman agent. *Cognitive Development, 16,* 637–656. doi:10.1016/S0885-2014(01)00043-0

Kotovsky, L., & Baillargeon, R. (1994). Calibration-based reasoning about collision events in 11-month-old infants. *Cognition, 51,* 107–129. doi:10.1016/0010-0277(94)90012-4

Kovack-Lesh, K. A., & Oakes, L. M. (2007). Hold your horses: How exposure to different items influences infant categorization. *Journal of Experimental Child Psychology, 98,* 69–93. doi:10.1016/j.jecp.2007.05.001

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience, 5,* 831–843. doi:10.1038/nrn1533

Kushnir, T., & Gopnik, A. (2007). Conditional probability versus spatial contiguity in causal learning: Preschoolers use new contingency evidence to overcome prior spatial assumptions. *Developmental Psychology, 43,* 186–196. doi:10.1037/0012-1649.43.1.186

Leslie, A. M. (1984a). Infant perception of a manual pick-up event. *British Journal of Developmental Psychology, 2,* 19–32. doi:10.1111/j.2044-835X.1984.tb00531.x

Leslie, A. M. (1984b). Spatiotemporal continuity and the perception of causality in infants. *Perception, 13,* 287–305. doi:10.1068/p130287

Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition, 25,* 265–288. doi:10.1016/S0010-0277(87)80006-9

Lyons, D. E., Young, A. G., & Keil, F. C. (2007). The hidden structure of overimitation. *Proceedings of the National Academy of Sciences of the United States of America, 104,* 19751–19756. doi:10.1073/pnas.0704452104

McGuigan, N., Whiten, A., Flynn, E., & Horner, V. (2007). Imitation of causally opaque versus causally transparent tool use by 3- and 5-year-old children. *Cognitive Development, 22,* 353–364. doi:10.1016/j.cogdev.2007.01.001

Meltzoff, A. N. (1988). Infant imitation after a 1-week delay: Long-term memory for novel acts and multiple stimuli. *Developmental Psychology, 24,* 470–476. doi:10.1037/0012-1649.24.4.470

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology, 31,* 838–850. doi:10.1037/0012-1649.31.5.838

Meltzoff, A. N. (2007a). Infants' causal learning: Intervention, observation, imitation. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 37–47). New York, NY: Oxford University Press.

Meltzoff, A. N. (2007b). The "like me" framework for recognizing and becoming an intentional agent. *Acta Psychologica, 124,* 26–43. doi:10.1016/j.actpsy.2006.09.005

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

Meltzoff, A. N., Kuhl, P. K., Movellan, J., & Sejnowski, T. J. (2009). Foundations for a new science of learning. *Science, 325,* 284–288. doi:10.1126/science.1175626

Michotte, A. (1962). *Causalité, permanence, et réalité phénoménales: Études de psychologie expérimentale* [Causality, permanence, and phenomenal reality: Studies of experimental psychology]. Louvain, Belgium: Publications Universitaires.

Muentener, P., & Carey, S. (2010). Infants' causal representations of state change events. *Cognitive Psychology, 61,* 63–86. doi:10.1016/j.cogpsych.2010.02.001

Nazzi, T., & Gopnik, A. (2000). A shift in children's use of perceptual and causal cues to categorization. *Developmental Science, 3,* 389–396. doi:10.1111/1467-7687.00133

Nielsen, M., & Tomaselli, K. (2010). Overimitation in Kalahari bushmen children and the origins of human cultural cognition. *Psychological Science, 21,* 729–736. doi:10.1177/0956797610368808

Oakes, L. M., & Cohen, L. B. (1990). Infant perception of a causal event. *Cognitive Development, 5,* 193–207. doi:10.1016/0885-2014(90)90026-P

Pearl, J. (2000). *Causality: Models, reasoning, and inference.* New York, NY: Cambridge University Press.

Phillips, A. T., & Wellman, H. M. (2005). Infants' understanding of object-directed action. *Cognition, 98,* 137–155. doi:10.1016/j.cognition.2004.11.005

Piaget, J. (1930). *The child's conception of physical causality.* New York, NY: Harcourt Brace.

Piaget, J. (1954). *The construction of reality in the child* (M. Cook, Trans.). New York, NY: Basic Books. doi:10.1037/11168-000

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach.* Cambridge, MA: MIT Press.

Rogoff, B., Paradise, R., Arauz, R. M., Correa-Chávez, M., & Angelillo, C. (2003). Firsthand learning through intent participation. *Annual Review of Psychology, 54,* 175–203. doi:10.1146/annurev.psych.54.101601.145118

Rovee-Collier, C. (1987). Learning and memory in infancy. In J. D.

Osofsky (Ed.), *Handbook of infant development* (2nd ed., pp. 139–168). Oxford, England: Wiley.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274,* 1926–1928. doi:10.1126/science.274.5294.1926

Saxe, R., Tenenbaum, J. B., & Carey, S. (2005). Secret agents: Inferences about hidden causes by 10- and 12-month-old infants. *Psychological Science, 16,* 995–1001. doi:10.1111/j.1467-9280.2005.01649.x

Saxe, R., Tzelnic, T., & Carey, S. (2007). Knowing who dunnit: Infants identify the causal agent in an unseen causal interaction. *Developmental Psychology, 43,* 149–158. doi:10.1037/0012-1649.43.1.149

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences, 4,* 299–309. doi:10.1016/S1364-6613(00)01506-0

Schulz, L. E., & Bonawitz, E. B. (2007). Serious fun: Preschoolers engage in more exploratory play when evidence is confounded. *Developmental Psychology, 43,* 1045–1050. doi:10.1037/0012-1649.43.4.1045

Schulz, L. E., Gopnik, A., & Glymour, C. (2007). Preschool children learn about causal structure from conditional interventions. *Developmental Science, 10,* 322–332. doi:10.1111/j.1467-7687.2007.00587.x

Sobel, D. M., & Kirkham, N. Z. (2006). Blickets and babies: The development of causal reasoning in toddlers and infants. *Developmental Psychology, 42,* 1103–1115. doi:10.1037/0012-1649.42.6.1103

Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2004). Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cognitive Science, 28,* 303–333. doi:10.1016/j.cogsci.2003.11.001

Spirtes, P., Glymour, C., & Scheines, R. (1993). *Causation, prediction, and search (Lecture notes in statistics).* New York, NY: Springer-Verlag.

Tomasello, M. (1999). *The cultural origins of human cognition.* Cambridge, MA: Harvard University 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* (pp. 77–85). Hove, England: Erlbaum.

Wellman, H. M., & Gelman, S. A. (1992). Cognitive development: Foundational theories of core domains. *Annual Review of Psychology, 43,* 337–375. doi:10.1146/annurev.ps.43.020192.002005

Williamson, R. A., Jaswal, V. K., & Meltzoff, A. N. (2010). Learning the rules: Observation and imitation of a sorting strategy by 36-month-old children. *Developmental Psychology, 46,* 57–65. doi:10.1037/a0017473

Williamson, R. A., Meltzoff, A. N., & Markman, E. M. (2008). Prior experiences and perceived efficacy influence 3-year-olds' imitation. *Developmental Psychology, 44,* 275–285. doi:10.1037/0012-1649.44.1.275

Woodward, J. (2003). *Making things happen: A theory of causal explanation.* New York, NY: Oxford University Press.

Woodward, J. (2007). Interventionist theories of causation in psychological perspective. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 19–36). New York, NY: Oxford University Press.

Received July 22, 2010
Revision received November 17, 2011
Accepted December 2, 2011 ■