Considering Counterfactuals

The Relationship between Causal Learning and Pretend Play

Alison Gopnik* and Caren M. Walker*

Many researchers have long assumed imaginative play critical to the healthy cognitive, social, and emotional development of children, which has important implications for early-education policy and practice. But, the authors find, a careful review of the existing literature highlights a need for a better theory to clarify the nature of the relationship between pretend play and childhood development. In particular, they ask why children spend so much time engaging in unreal scenarios at a time when they know relatively little about the real world? The authors review the idea that children pretend because it exercises their developing ability to reason counterfactually – an ability essential for causal reasoning and learning. The authors present a look at their study in progress aimed at assessing their theory. According to the model of play they outline, imaginative play serves as an engine of learning. Such play arises out of the human capacity for causal cognition and feeds back to help develop causal reasoning skills. Key words: Bayesian learning methods; causal learning; counterfactual reasoning; pretend play; probabilistic models

American Journal of Play, volume 6, number 1 © The Strong

*Joint first authors

Across species, the activities typically involved in play are those that will become important in adulthood (Bekoff and Byers 1998). Play, then, is a type of exploratory learning, in which the young animal engages in a variety of behaviors in a low-risk, low-cost context. Over the last ten years or so, a growing body of research has focused on a unique form of exploratory play in human children, involving a process of informal experimentation on the world. This previous work demonstrates that children’s early exploration of the external world during free play helps them learn the complexities of causal relationships (e.g., see Schulz 2012 for a review).
However, human children also engage in a particularly distinctive kind of internal exploratory play—pretend play. During pretense, children not only play with objects and social partners in the actual world, they also construct (sometimes rather elaborate) unreal scenarios about possible worlds. The ideas we address in this article relate to a long-standing question in developmental psychology about the apparently paradoxical nature of this type of pretend play: Why do young children spend so much of their time and resources on unreality, given that they are still learning about the real world?

Classically, researchers (Freud 1922; Piaget 1962) offered a rather uncharitable characterization of pretend play—attributing its prevalence in childhood to children’s inability to differentiate between fantasy and reality. However, decades of research have shown that, to the contrary, children seem quite proficient at distinguishing between the two (Flavell, Flavell, and Green 1987; Morison and Gardner 1978; Skolnick and Bloom 2006; Taylor 1999; Walton 1990; Woolley and Cox 2007; Woolley and Ghossainy 2013). Many of these early deflationary theories of pretense in childhood also ignored some of the most prominent features of pretend play, features that suggest it is the result of cognitive competence rather than cognitive limitations. Pretense is unique to human beings; it is often social in nature; and it becomes increasingly elaborate over the course of early childhood (Harris 2000). Indeed, one of the more compelling features of pretend play is its continuity throughout the lifespan: Adults spend large portions of their lives engaged with fictional worlds.

No wonder that developmental psychologists and educators have long suspected pretend play contributes to learning. Historically however, there have been surprisingly few empirical studies that provide strong support for this idea (see Lillard et al. 2013 for a recent review). We have argued elsewhere (Walker and Gopnik 2013a) that the lack of a unifying theory of pretend
play has been detrimental to its empirical study because a coherent theory serves to generate a set of testable predictions to guide the direction of research. Even some of the more theoretically sophisticated accounts of pretense tend to interpret pretend play as an epiphenomenon or a byproduct of other abilities (e.g., children’s developing ability to reason about the minds of others) rather than as an activity that could itself shape learning (Baron-Cohen 1995; Currie 1995; Leslie 1987; Lillard 2001; Nichols and Stich 2000).

Instead, we have suggested that the type of imagination-based thought that appears during pretend play is continuous with and deeply connected to reality-directed thinking (Buchsbaum et al. 2012; Gopnik 2009; Skolnick and Gopnik 2013; Walker and Gopnik 2013b). In particular, we have argued that the ability to represent and reason about causal relations may underlie the ability to imagine possible worlds, and further, that the ability to imagine alternative possibilities may feed back to aid in the development of causal reasoning and learning (Walker and Gopnik 2013a).

These ideas come from a broader proposal in developmental psychology called the “theory theory” (Carey 1985; Gopnik and Meltzoff 1997; Wellman 1990; Wellman and Gelman 1998), which views the developing child as a young scientist, acquiring and revising causal theories—abstract, coherent representations of the cause and effect relationships in the world (see Gopnik and Wellman 2012 for a review). Theory theorists believe the process of causal theory revision analogous to the process of scientific theory change. The child begins with a currently held theory, then observes some new evidence that conflicts with this theory, then forms a range of alternative hypotheses to test against the evidence, and eventually revises her theory to better fit the evidence observed in the world. Importantly, theory theorists assert, this
process of theory change is typically facilitated by the child’s own exploration of the world, often through play.

To illustrate, consider a young child who holds an incorrect causal theory of flotation. Because children’s intuitions often fail to differentiate between weight and density (Carey 2009), this child believes that heavy objects tend to sink and light objects tend to float. While in the bathtub one evening, she observes some new evidence that conflicts with her currently held theory: a small marble sinks to the bottom of the tub, while a much heavier toy—perhaps a plastic truck—floats on the surface of the water. In response to this anomalous evidence, she may form a range of alternative hypotheses to test against the evidence. For example, perhaps this toy is specially equipped with a motor to keep it afloat. However, after a variety of investigations—including an exploration of the truck, observations of other objects in the tub, or testimony from a knowledgeable adult that conflicts with her theory of floatation—she may eventually revise her theory to better fit the actual causal structure of the world.

Although these ideas about theory change have been around for some time, researchers have historically failed to provide a precise account of the cognitive mechanisms that underlie the process of theory change. Recently however, developmental psychologists in collaboration with computer scientists and philosophers have begun building a computational theory that describes these representations and learning mechanisms. This has been a central part of a larger movement in cognitive science. “Probabilistic models” and “Bayesian learning methods,” two of the key ideas in this movement, have been applied to a broad set of problems in human cognition. These include problems central to cognitive development—in particular, how we derive rich, abstract representations of the world from the limited concrete data available.
These computational models define learning as the process of building structured, abstract representations of how the world works. The models are probabilistic because such learning involves assessing the probability that each possible representation accurately portrays the causal structure of the world. A full exposition of these ideas may be found elsewhere (e.g., for reviews see Gopnik 2012; Gopnik and Wellman 2012). The three main concepts we summarize below, however, are particularly relevant to the relationship between the mechanisms underlying causal learning and pretend play.

**Concept 1: Causal Models and Probabilistic Inference**

One of the central ideas of this framework suggests that children’s complex, coherent representations of causal relationships may be expressed in a kind of causal map, or abstract picture of how the world works. In many ways, these causal maps are analogous to the more familiar spatial maps that depict the various locations of objects in relation to one another. Having a spatial map is useful because it provides a nonegocentric representation of the spatial relations among objects. In turn, this allows the construction of a variety of possible routes to navigate in space.

Similarly, having a causal map provides a complex representation of causal relationships—a picture of how one thing is causally connected to another—that enables the learner to generate inferences, make predictions and plan ahead, consider the consequences of possible events, and perform effective actions in the world. These causal maps may be formalized in graphical descriptions called, causal models or “Bayes nets” (Pearl 2000; Spirites,
Glymour, and Scheines 1993), which mathematically represent the causal relationships between objects and events.

Each causal model may then be understood as a representation of a particular hypothesis about the network of causal relationships—or causal structure—that exists in the world. If the hypothesis that the learner currently holds about the actual causal structure of the world proves correct, then the predictions generated by the causal model will turn out to be accurate. However, if the hypothesis that the learner currently holds is incorrect, then the causal model will fail to predict her observations accurately. To return to our previous example, if a child believes that the weight of an object causes it to sink or float, then her current causal model fails to account for her observation of the sinking marble and floating truck. This failure should prompt her, the learner, to adjust her causal model to better approximate the structure of the actual world (in other words, she undergoes a process of theory change). Therefore, these models can describe representations of the world and explain how these representations enable her to make a wide range of new predictions. Critically, however, the systematic link between causal structure and evidence also allows her to reverse the process and to make inferences about the nature of the causal structure from the evidence generated by the model. It lets her decide which causal model best accounts for the evidence and so leads her to adopt the most likely hypothesis.

The intellectual advance that turned these causal models into one of the most powerful tools used by statisticians and computer scientists came with the integration of ideas about probability into this basic inferential framework. A great many hypotheses are, in principle, compatible with any pattern of evidence, so how can we decide on the best one? Although many hypotheses may be compatible with the evidence, some hypotheses will be more likely to have generated the evidence than others. In other words, rather than simply generating a “true” or
“false” inference about whether a particular hypothesis is true with respect to the world, the learner considers multiple hypotheses (i.e., multiple causal models) and assigns probabilities to those hypotheses (see Perfors et al. 2011 for a general introduction). The integration of probability also gives this computational approach the combination of stability and flexibility that characterizes early learning and inductive inference. A learner will be reluctant to give up a strongly confirmed hypothesis, but even the most entrenched idea can be rejected if enough counterevidence accumulates.

Although probabilistic models were originally designed to be ideal rational accounts of how a scientist or a computer could best solve a learning problem, ten years of empirical research have shown that these computational accounts can also effectively explain very young children’s learning in a wide array of domains. For example, we can manipulate the evidence preschool children (and even infants) observe about a causal system and thus their beliefs about the probability of various hypotheses about how the system works. Then we can examine what conclusions they draw. Even very young children characteristically choose the hypotheses with the greatest probability (Gopnik et al. 2001; Gopnik et al. 2004; Gweon and Schulz 2011; Kushnir and Gopnik 2005, 2007; Schulz et al. 2007; Sobel and Kirkham 2006; Sobel et al. 2004).

**Concept 2: Causal Relations Imply Counterfactuals**

The second idea contends that every single causal relationship may be defined in terms of a counterfactual: The claim that X causes Y means that if you changed X then it would lead to a change in Y (Lewis 1986). This idea distinguishes causation from mere correlation—the fact that these events are yoked to one another. For example, you might see a correlation between smoking and both yellow nicotine-stained fingers and lung cancer and that might lead you to
predict that someone who smokes or someone who has yellow fingers will be more likely to get cancer. But the counterfactuals are different. If the person had not smoked, they would not have gotten cancer. But they might not have had yellow fingers (they might have always washed them), and cancer would still have developed.

This ability to reason about the causal dependence between events enables the learner to generate predictions and facilitate planning (e.g., “what would happen if I were to do X?”) and to reason about counterfactuals or alternatives (e.g., “what would have happened if I had done X?”). According to this picture, the ability to imagine alternative ways the world could be proves absolutely central to this representation of our causal knowledge. Each relationship between each of the events in a causal model actively generates a set of possible worlds, some of which are factual and some of which are counterfactual.

Counterfactual reasoning is also fundamental to the very process of learning in a probabilistic framework. In this learning process, the learner makes predictions about the evidence based on his current highest-probability hypothesis and compares them to the predictions made by a set of lower-probability hypotheses. In other words, during learning, the learner takes whatever causal structure he currently believes to be true (or his current hypothesis), and examines the evidence to assess the probability that such a causal structure best fits the world compared to other possible (counterfactual) causal structures.

**Concept 3: The Power of Interventions for Learning**

Finally, one of the main advantages of applying the causal-models framework is that it includes a means for representing “interventions,” or possible actions on a causal system. A causal model enables us to predict not just what will happen but what would happen if we made things
different. Interventions performed on causal models change the representation of the causal structure by cutting off an event from its usual causes and observing the effects of this change on the probability of the other connected events. This is precisely what we do in science. For example, if we want to examine the relationship between smoking and cancer, we intervene on one variable (e.g., reduce smoking) and observe the effects on the other (e.g., instances of cancer). Importantly, this procedure may be used to represent actual interventions in the physical world, and exactly the same procedure may also be used to generate pretend or hypothetical interventions in the imagination (Gopnik 2009; Sloman 2005; Pearl 2000).

The ability to imagine possible interventions gives us a powerful tool with which to plan our actions in advance. When we do so, we make an assumption that we know is false with respect to the current state of the world and then follow the various implications downstream. This, of course, is also what we do in pretense and in fiction: A play partner or an author makes a particular assumption about the world (e.g., time travel exists; your toys come alive when you leave the room), and follows the causal implications of this assumption throughout the narrative. We therefore propose that these crucially important abilities for learning about the actual world (creating possible causal interventions and testing alternative causal hypotheses) depend on the same cognitive machinery that children use when they pretend (adopting a premise that is currently not true, creating a sequence of events that follows from the premise, and quarantining the result of this process from reality).

To return to the paradox of pretend play, we believe that pretense is simply a precocious display of children’s developing abilities to engage in counterfactual reasoning about causal relationships. Indeed, Harris and colleagues (Harris 2000; Harris, Kavanaugh, and Dowson 1997) have pointed out that pretend play draws heavily from children’s existing causal
knowledge about the real world. Pretense involves generating a novel premise and reasoning through the causal consequences.

For example, when participating in pretense, children routinely engage with premises that contradict their own knowledge. They ignore the fact that the teacup is empty and proceed to wipe up the imagined tea when the cup is overturned. Indeed, anytime children act on the outcome of a pretend transformation, they are necessarily setting aside their interpretation of the real world and reasoning about the causal consequences of a particular premise. In fact, children are quite good at reasoning about causal transformations in a pretend context (Harris, Kavanaugh, and Dowson 1997). Moreover, children are actually better able to engage in cognitively demanding deductive reasoning tasks when they are presented in pretend contexts, particularly when the premises are false with respect to the world (e.g., “All fish fly. Dot is fish. Does Dot fly?”) (Dias and Harris 1990).

In our lab, we have begun generating evidence for the relationship between counterfactual reasoning and pretend play in preschool-aged children. For example, in a first set of experiments (Buchsbaum et al. 2012), we introduced three- and four-year-old children to a causal toy, called a “blicket detector,” that activates and plays music when certain objects (blickets) are placed on top of it. We asked them to make counterfactual inferences about the machine. “What would happen,” we ask, “if this block (which is not a blicket) were a blicket?” We also prompted them to engage in pretend play about the machine with a set of novel objects. “Pretend this box is the blicket detector,” we said, “and pretend this block is a blicket, what should we pretend next?” Findings from this study demonstrate a significant and specific correlation between children’s pretend play and their ability to generate counterfactuals, even controlling for general cognitive abilities. In particular, children who were unable to reason
about causal relationships in the pretend scenario were also unable to reason about the effect of counterfactual alternatives with the actual toy (e.g., “What if this block were not a blicket?”).

Of course, correlation does not imply causation. Pretend play might simply be a byproduct of the ability to generate alternative hypotheses. Therefore, to follow up these initial findings, we are currently running a series of studies in our lab. Across studies, we focus on the period of early learning during the preschool years, between three and five years of age. This is the time during development when pretend play is most prominent (Singer and Singer 1990). It is also the time when children are developing causal theories about the world (Carey 1985). First, we are currently examining whether children are able to reason about interventions on a complex causal system in a pretend scenario and also whether pretense (in particular) might support children’s causal reasoning (Walker, Buchsbaum, and Gopnik 2013).

To conduct these studies, we used a paradigm similar to one developed earlier in our lab by Laura Schulz and her colleagues (2007). Preschoolers were taught causal relationships using gear toys (a box with two interlocking gears and a switch) like those pictured in figure 1. Children first observe that flipping the switch causes both gears to spin. However, the toy might have at least three different causal structures—a chain (the switch causes gear A to spin, and gear A causes gear B to spin), a reverse chain (the switch causes gear B to spin, and gear B causes gear A to spin), or a common effect structure (the switch causes gears A and B to spin independently). Each of these structures generates different patterns of evidence. For example, in a chain, if you remove gear A and flip the switch, gear B will not spin. Schulz and her coauthors demonstrated that given a particular pattern of evidence, four-year-olds chose the right causal structure and vice versa – given the structure, children predicted the evidence.
In our current study, four- and five-year-old children were initially taught the “real” causal structure of the gear toy and the consequences of various interventions. After these children demonstrated they had learned this causal structure, we asked them to pretend that the toy worked in a different way. So, for example, if we had taught them that the toy worked as a causal chain (the switch causes gear A to spin, and gear A causes gear B to spin), we asked them to pretend that the toy worked as a reverse causal chain (the switch causes gear B to spin, and gear B causes gear A to spin). We then asked them to reason about a series of interventions, assuming the pretend structure. For example, we removed gear A from the toy, we pretended to flip the switch, and we asked them what they pretended happened to gear B. After this first trial, we introduced a second gear toy, with a distinct causal structure, and repeated the entire procedure.

By introducing children to the gear toy, we were able to examine their counterfactual reasoning abilities and their intuitions about the effects of pretend actions on a completely novel complex causal system. Unlike previous research examining children’s ability to reason about the outcome of pretend transformations (e.g., Harris, Kavanaugh, and Dowson 1997), this paradigm prevented children from answering our questions from familiar scripts and general knowledge. The gear toy also provided a flexible, multivariable causal structure, which enabled us to manipulate the evidence that children observed and examine the consequences on their reasoning.

The results of this study indicate that children can, and do, in fact, reason about the outcomes of interventions to complex causal structures in the context of a pretend scenario. In fact, children’s inferences in this study were surprisingly accurate and precisely as accurate as the predictions about real scenarios in the original study (Schulz et al. 2007). We are currently
running an additional study to compare children’s performances in this pretend context (e.g., “Let’s pretend the toy worked a different way.”) to their performances in a standard hypothetical context (e.g. “What if the toy worked a different way?”). We predict that children’s performances across studies will be highly correlated. It is also possible that children may actually do better in the pretense condition. They may be less able to explore alternatives accurately in the less playful context, especially given the existing literature on children’s difficulties with counterfactual reasoning.

This prediction is consistent with much of the earlier work conducted by Paul Harris and others. They demonstrated that pretend scenarios can serve as a cognitive tool to foster abstract reasoning by prompting the learner to attend to the premises and treat them as quarantined from prior knowledge (Dias and Harris 1988, 1990; Harris and Leevers 2000; Hawkins et al. 1984). Additional support for this idea also comes from related work by Amsel, Trionfi, and Campbell (2005), who found that reasoning about make-believe suppositions and reasoning from hypothetical suppositions is highly correlated in older children and adults.

From a practical perspective this theory and set of results supports the idea that pretend play helps children learn. This is of course, an idea that has seemed intuitive to generations of early-childhood educators. However, it is now under increasing pressure from both parents and policy makers who endorse a more academic model of early childhood education. As Lillard et al. (2013) have effectively demonstrated, however, there is surprisingly little strong empirical evidence for the intuitive claim. We think this is due to a paucity of good recent research and also to the fact that earlier research has used a very general and broad definition of both play and learning (Walker and Gopnik 2013a).
From a theoretical perspective, there is little reason to expect that all types of play—from exploratory “getting into everything,” to pretend, to rough and tumble play, to playground games with rules—would be related to all types of learning, from reasoning about mental states to math. Instead we propose that pretend play, in particular, is related to a very specific but very important type of learning and reasoning—namely the kind of counterfactual reasoning that is intimately involved with causal knowledge and learning. We hope that our research will be one among many working to elucidate the relationship between play and learning in all its complexity and promise.

REFERENCES


Bekoff, Marc, and John A. Byers, eds. 1998. *Animal Play: Evolutionary, Comparative, and Ecological Perspectives.*


Freud, Sigmund. 1922. *Beyond the Pleasure Principle.* Translated by C. J. M. Hubback.


